

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа Инженерная школа природных ресурсов
Направление подготовки 21.04.01 Нефтегазовое дело
Отделение школы (НОЦ) ОНД

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы	
Выделение продуктивных интервалов и прогноз промышленных характеристик в скважинах с применением алгоритмов машинного обучения на примере Шингинского нефтяного месторождения	

УДК 622.323:553.982.044(571.16)

Студент

Группа	ФИО	Подпись	Дата
2ТМ61	Балашов Дмитрий Александрович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ведущий эксперт ЦППС НД ТПУ	Меркулов В.П.	К.Г.-М.Н.		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор	Белозеров В.Б.	Д.Г.-М.Н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ведущий эксперт ЦППС НД ТПУ	Меркулов В.П.	К.Г.-М.Н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Ведущий эксперт ЦППС НД ТПУ	Чернова О.С.	К.Г.-М.Н.		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ООП

Код	Результат обучения
P1	Применять естественнонаучные, математические, гуманитарные, экономические, инженерные, технические и глубокие профессиональные знания в области современных нефтегазовых технологий для решения прикладных междисциплинарных задач и инженерных проблем, соответствующих профилю подготовки
P2	Планировать и проводить аналитические и экспериментальные исследования с использованием новейших достижений науки и техники, уметь критически оценивать результаты и делать выводы, полученные в сложных и неопределённых условиях; использовать принципы изобретательства, правовые основы в области интеллектуальной собственности
P3	Проявлять профессиональную осведомленность о передовых знаниях и открытиях в области нефтегазовых технологий с учетом передового отечественного и зарубежного опыта; использовать инновационный подход при разработке новых идей и методов проектирования объектов нефтегазового комплекса для решения инженерных задач развития нефтегазовых технологий, модернизации и усовершенствования нефтегазового производства
P4	Выбирать оптимальные решения в многофакторных ситуациях, владеть методами и средствами технического моделирования производственных процессов и объектов нефтегазовой отрасли; управлять технологическими процессами, обслуживать оборудование, использовать любой имеющийся арсенал технических средств, обеспечивать высокую эффективность при разработке нефтегазовых объектов
P5	Самостоятельно учиться и непрерывно повышать квалификацию в течение всего периода профессиональной деятельности; активно владеть иностранным языком на уровне, позволяющем работать в интернациональной среде, разрабатывать документацию и защищать результаты инженерной деятельности
P6	Работать эффективно в качестве члена и руководителя команды, умение формировать задания и оперативные планы всех видов деятельности, распределять обязанности членов команды, нести ответственность за результаты работы; координировать работу групп по извлечению и совершенствованию добычи нефти, газа и газового конденсата, передавать знания через наставничество и консультирование

P7	Внедрять, эксплуатировать и обслуживать современные машины и механизмы для реализации технологических процессов нефтегазовой области, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья и безопасности труда, выполнять требования по защите окружающей среды
P8	Предлагает процедуры анализа и оценки геологических объектов по данным геолого-геофизической изученности и скважинных исследований в целях разработки программы исследований и технологических мероприятий для конкретного типа месторождений углеводородов
P9	Предлагать процедуры оценки эффективности промысловых работ и оптимизации работы оборудования при добыче нефти, газа и газового конденсата, обеспечение энергоэффективности технологических процессов
P10	Обеспечивать внедрение новых методов при построении геолого-геофизических моделей; материалов и нефтегазового оборудования в осложненных условиях эксплуатации нефтяных и газовых скважин
P11	Контролировать выполнение требований регламентов для обеспечения добычи нефти, газа и газового конденсата
P12	Совершенствовать, разрабатывать мероприятия и/или подготавливать бизнес-предложения по технологическому процессу и технологическим мероприятиям при добыче нефти, газа и газового конденсата
P13	Корректировать программы работ по добыче нефти, газа и газового конденсата, выбирать и принимать решения в нестандартных ситуациях, опираясь на государственные стандарты в области нефтегазодобычи

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа Инженерная школа природных ресурсов
Направление подготовки (специальность) 21.04.01 Нефтегазовое дело
Отделение школы (НОЦ) ОНД

УТВЕРЖДАЮ:
Руководитель ООП

(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации
(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
2ТМ61	Балашов Дмитрий Александрович

Тема работы:

Выделение продуктивных интервалов и прогноз промышленных характеристик в скважинах с применением алгоритмов машинного обучения на примере Шингинского нефтяного месторождения	
Утверждена приказом директора (дата, номер)	08.08.2018, 7260/с

Срок сдачи студентом выполненной работы:

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе</p> <p><i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<p>Геофизические исследования скважин месторождения, результаты интерпретации ГИС, параметры добычи скважин, информация о пластовом и забойном давлениях, информация о геологическом строении Западно-Сибирской нефтегазоносной провинции в районе месторождения.</p>
<p>Перечень подлежащих исследованию, проектированию и разработке вопросов</p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов,</i></p>	<p>Основные сведения о месторождении, геологическом строении и нефтегазоносности, обзор литературных источников связанных с тематикой работы, разработка основной методики выделения продуктивных интервалов в рамках кластерного анализа, применение результатов кластерного анализа для прогноза промышленных скважинных характеристик, экономическая оценка</p>

подлежащих разработке; заключение по работе).	перспективности методики, социальная ответственность.
Перечень графического материала <i>(с точным указанием обязательных чертежей)</i>	1. Обзорная карта территории 2. Тектоническая карта структурного яруса 3. Нефтегазоносность Западно-Сибирской нефтегазоносной провинции 4. Структурная карта пласта Ю ₁ ¹ 5. Итеративный процесс оценки кластеризации разработанной метрикой 6. Оптимизация кластерного анализа 7. Дендрограмма 8. Планшеты скважин с результатами кластерного анализа 9. Карты толщин кластеров 10. Матрица корреляции 11. Результаты регрессионного анализа 12. Результаты анализа кривых темпов падения 13. Профили добычи проектных скважин
Консультанты по разделам выпускной квалификационной работы <i>(с указанием разделов)</i>	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Профессор, д.г.-м.н. Белозеров В.Б.
Социальная ответственность	Ведущий эксперт ЦППС НД ТПУ, к.г.-м.н. Меркулов В.П.
Названия разделов, которые должны быть написаны на русском и иностранном языках:	
4 Разработка методики	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	27.06.2018
--	------------

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ведущий эксперт ЦППС НД ТПУ	Чернова О.С.	К.г.-м.н.		27.06.2018

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
2ТМ61	Балашов Дмитрий Александрович		27.06.2018

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
2ТМ61	Балашов Дмитрий Александрович

Школа		Отделение школы (НОЦ)	
Уровень образования	магистратура	Направление/специальность	21.04.01. нефтегазовое дело

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Экономические параметры, являющиеся актуальными на период 2018-2020 годов, цены на капитальные и операционные затраты
2. Нормы и нормативы расходования ресурсов	Ставки заработных плат, нормы отчислений в амортизационный фонд, нормы на расходы материалов и инструмента
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Применяется консолидированная система налогообложения, включающая НДС (18%), НДС/П (расчетная величина), налог на прибыль (20%), экспортную пошлину (132 \$/тонна)

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого и инновационного потенциала НТИ	Технико-экономическое обоснование возможности бурения скважин
2. Разработка устава научно-технического проекта	Расчет капитальных и эксплуатационных затрат, налоговых отчислений
3. Определение ресурсной, финансовой, экономической эффективности	Обоснование целесообразности бурения скважин

Перечень графического материала (с точным указанием обязательных чертежей):

1. Профиль ЧДД по месяцам
2. Расчет ЧДД на примере одной скважины
3. Структура операционных затрат и системы налогообложения

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор	Белозеров В.Б.	д.г.-м.н.		27.06.2018

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
2ТМ61	Балашов Дмитрий Александрович		27.06.2018

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
2ТМ61	Балашов Дмитрий Александрович

Школа		Отделение школы (НОЦ)	
Уровень образования	магистратура	Направление/специальность	21.04.01 Нефтегазовое дело

Исходные данные к разделу «Социальная ответственность»:

<p>1. Описание рабочего места (рабочей зоны, технологического процесса, механического оборудования) на предмет возникновения:</p> <ul style="list-style-type: none"> – вредных проявлений факторов производственной среды (метеоусловия, вредные вещества, освещение, шумы, вибрации, электромагнитные поля, ионизирующие излучения) – опасных проявлений факторов производственной среды (механической природы, термического характера, электрической, пожарной и взрывной природы) – негативного воздействия на окружающую природную среду (атмосферу, гидросферу, литосферу) – чрезвычайных ситуаций (техногенного, стихийного, экологического и социального характера) 	<p>Описать возможные вредные факторы, влияющие на окружающую среду (биосферу, гидросферу, атмосферу и литосферу). Спланировать меры по предупреждению возможных аварийных ситуаций в процессе бурения и эксплуатации скважин.</p>
<p>2. Перечень законодательных и нормативных документов по теме</p>	<p>1. ФЗ «Об охране окружающей среды», 2. Приказ «Об утверждении Федеральных норм и правил в области промышленной безопасности»</p>

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<p>1. Анализ выявленных вредных факторов проектируемой производственной среды в следующей последовательности:</p> <ul style="list-style-type: none"> – физико-химическая природа вредности, её связь с разрабатываемой темой; – действие фактора на организм человека; – приведение допустимых норм с необходимой размерностью (со ссылкой на соответствующий нормативно-технический документ); – предлагаемые средства защиты (сначала коллективной защиты, затем – индивидуальные защитные средства) 	<p>Предлагается контроль освещения и температурных показателей на рабочем месте в соответствии с нормативными положениями, установка ограждающих приспособлений</p>
<p>2. Анализ выявленных опасных факторов проектируемой производственной среды в следующей последовательности</p> <ul style="list-style-type: none"> – механические опасности (источники, средства защиты); – термические опасности (источники, средства защиты); – электробезопасность (в т.ч. статическое электричество, молниезащита – источники, средства защиты); – пожаровзрывобезопасность (причины, профилактические мероприятия, первичные средства пожаротушения) 	<p>Необходимо изучить такие вредные факторы, как недостаточная освещенность, электрические замыкания. Требуется снизить пожаровзрывоопасность, возможность утечек газа. Исключить нанесение вреда здоровью в результате открытого контакта с оборудованием.</p>
<p>3. Охрана окружающей среды:</p> <ul style="list-style-type: none"> – защита селитебной зоны – анализ воздействия объекта на атмосферу (выбросы); – анализ воздействия объекта на гидросферу (сбросы); 	<p>Во время бурения скважин возможны ГНВП, также негерметичность амбара может привести к разливу бурового раствора, разливы нефти при капитальном ремонте скважины</p>

<ul style="list-style-type: none"> – анализ воздействия объекта на литосферу (отходы); – разработать решения по обеспечению экологической безопасности со ссылками на НТД по охране окружающей среды. 	
4. Защита в чрезвычайных ситуациях: <ul style="list-style-type: none"> – перечень возможных ЧС на объекте; – выбор наиболее типичной ЧС; – разработка превентивных мер по предупреждению ЧС; – разработка мер по повышению устойчивости объекта к данной ЧС; – разработка действий в результате возникшей ЧС и мер по ликвидации её последствий 	<p><i>Возможные ЧС: обрыв электрического кабеля большого напряжения, проявление ГНВП во время бурения, производственные аварии, природные ЧС, такие как сильный мороз.</i></p>
5. Правовые и организационные вопросы обеспечения безопасности: <ul style="list-style-type: none"> – специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны 	<p><i>Место работы должно быть оборудовано в соответствии с ГОСТ 12.2.032-78</i></p>
Перечень графического материала:	
<i>При необходимости представить эскизные графические материалы к расчётному заданию (обязательно для специалистов и магистров)</i>	

Дата выдачи задания для раздела по линейному графику	
---	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ведущий эксперт ЦППС НД ТПУ	Меркулов В.П.	к.г.-м.н.		27.06.2018

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
2ТМ61	Балашов Дмитрий Александрович		27.06.2018

РЕФЕРАТ

Выпускная квалификационная работа содержит 126 страниц, 46 рисунков, 9 таблиц, 21 источник, 3 приложения.

Ключевые слова: кластеризация, кластерный анализ, кластер, машинное обучение, регрессия, регрессионный анализ, линейная регрессия, кривые темпов падения, рейтинг бурения, уплотняющее бурение.

Объектом исследования являются геологические особенности месторождения, промышленные характеристики скважин, темп падения добычи на скважинах.

Цель работы – разработка методики выделения продуктивных интервалов в скважинах с помощью алгоритмов машинного обучения без учителя и увязка их с промышленными характеристиками скважин для прогноза последних с применением регрессионного анализа.

В процессе исследования проводились работы по изучению алгоритмов машинного обучения, как без учителя, так и с учителем, разработка подхода применения кластеризации для геологических задач, разработка подхода для предсказания запускных дебитов на скважинах с помощью регрессионного анализа, изучение поведения добычи на скважинах для получения усредненного темпа падения добычи нефти и жидкости.

В результате исследования была достигнута поставленная цель: была разработана методика, позволяющая как получать более совершенные результаты при работе с геологическими данными на стадии геологоразведочных работ, так и прогнозировать добычу на проектных скважин на разрабатываемых месторождениях.

Степень внедрения: методика может использоваться сотрудниками нефтяных компаний для широкого спектра, как геологических задач, так и задач, связанных с улучшением системы разработки месторождения.

Область применения: интерпретация ГИС, построение геологической модели, построение концептуальной модели, планирование уплотняющих скважин.

Экономическая эффективность/значимость работы: методика позволяет проводить более качественный анализ, что в результаты приводит к экономии денежных средств.

В будущем планируется автоматизировать методику для возможности ее использования специалистами, не обладающими глубокими знаниями программирования и машинного обучения.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

АК – акустический каротаж;

БКЗ – боковое каротажное зондирование;

ВИКИЗ – высокочастотное индукционное каротажное изопараметрическое зондирование;

ВНК – водонефтяной контакт;

ГГК-П – плотностной гамма-гамма каротаж;

ГИС – геофизические исследования скважин;

ГК – гамма каротаж;

ГРП – гидроразрыв пласта;

ДП – денежный поток;

ДСО – дисконтированный срок окупаемости;

ИК – индукционный каротаж;

МГРП – многостадийный гидроразрыв пласта;

НК – нейтронный каротаж;

РИГИС – результаты интерпретации геофизических исследований скважин;

СТК – спектральный гамма каротаж;

ЧДД – чистый дисконтированный доход;

DBSCAN – density-based spatial clustering of application with noise;

PE – индекс фотоэффекта;

SPE – общество инженеров нефтяников.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	13
1 СВЕДЕНИЯ О МЕСТОРОЖДЕНИИ	15
1.1 Общие сведения	15
1.2 Геолого-физическая изученность.....	18
1.3 Тектоническое строение территории	19
1.4 Стратиграфия месторождения	21
1.5 Нефтегазоносность	26
2 ЛИТЕРАТУРНЫЙ ОБЗОР	29
3 ОБЪЕКТ ИССЛЕДОВАНИЯ, ИСХОДНЫЕ ДАННЫЕ И СПОСОБ РЕАЛИЗАЦИИ	39
3.1 Объект исследования.....	39
3.2 Исходные данные.....	40
3.3 Способ реализации	40
4 РАЗРАБОТКА МЕТОДИКИ	42
4.1 Выбор скважин для оценки алгоритмов кластеризации	42
4.2 Выбор алгоритмов кластеризации	44
4.3 Подготовка данных.....	54
4.4 Разработка метрики оценка качества кластеризации	57
4.5 Машинное обучение с учителем для прогноза запускных дебитов	70
4.6 Метод кривых темпов падения добычи	78
4.7 Выбор скважин для уплотняющего бурения	81
4.8 Стандартный подход составления рейтинга бурения	83
5 РЕЗУЛЬТАТЫ РАЗРАБОТКИ.....	85
6 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ.....	86
6.1 Перспективность методики.....	86
6.2 Реализация технической задачи	87
6.3 Техничко-экономическое обоснование применимости методики	87
7 СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ	95

7.1	Аварийные ситуации в процессе бурения и добычи	98
7.2	Охрана окружающей среды	99
7.3	Ликвидация скважин	100
7.4	Безопасность и здоровье персонала	101
ЗАКЛЮЧЕНИЕ.....		102
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ		104
Приложение А.....		106
Приложение Б		107
Приложение В.....		108

ВВЕДЕНИЕ

Целью работы является разработка методики, которая позволяет выделять продуктивные интервалы в скважинах с помощью машинного обучения без учителя и в дальнейшем увязывать результаты кластеризации с промышленными характеристиками скважин с помощью регрессионного анализа.

Основными задачами, решение которых представлено в данной работе, являются:

1. ознакомление с геологическим строением выбранного для анализа месторождения, анализ предоставленных для работы данных ГИС, РИГИС и данных по добычи скважин;

2. разработка методологии обработки исходных данных для возможности их применения в машинном обучении, оптимизация методологии средствами языка программирования Python;

3. разработка алгоритма работы с машинным обучением без учителя для задач нефтяной индустрии, выбор и разработка оптимальных метрик оценки качества результатов кластеризации для геологических задач, обоснование процесса выбора алгоритма при кластерном анализе, проведение кластеризации по каротажным кривым для выделения продуктивных интервалов с помощью анализа выделенных коллекторских и неколлекторских кластеров, сравнение результатов разных алгоритмов и выбор наиболее оптимального;

4. разработка и применение подхода обработки результатов кластеризации для возможности использования их при регрессионном анализе;

5. обоснование и выбор регрессионной модели для прогноза запускных промышленных характеристик скважин, таких как запускные дебиты нефти и жидкости;

6. обоснование возможности применения уплотняющего бурения на месторождении, планирование проектных скважин и получение типов кластеров и их мощностей в скважинах с помощью методов интерполяции;

7. получение запускных дебитов на проектных скважинах путем применения полученной в процессе регрессионного анализа модели;

8. анализ кривых темпов падения добычи нефти и жидкости на месторождении с целью получения усредненных кривых падения добычи по месторождению для нефти и жидкости;

9. оценка полученной методики для составления рейтинга бурения уплотняющих скважин путем сравнения ее со стандартным подходом;

10. оценка экономической эффективности разработанной методики;

11. анализ возможных негативных последствий на окружающую среду в результате бурения уплотняющих скважин и планирование способов их минимизации, планирование мероприятий для обеспечения безопасной работы персонала.

Работа является актуальной ввиду того, что разработанный подход несет в себе не только теоретическое обоснование, но и возможность практического применения для разных задач нефтяной индустрии. Методикой могут пользоваться как геологи, так и нефтяники-разработчики. В работе доказана экономическая эффективность. Более того, методика использует последние наработки в области машинного обучения, а также и совершенно новые разработанные подходы, которые позволяют решать сложные задачи.

1 СВЕДЕНИЯ О МЕСТОРОЖДЕНИИ

1.1 Общие сведения

Шингинское нефтяное месторождение расположено в Западно – Сибирской нефтегазоносной провинции, которая находится на территории Российской Федерации. В административном отношении месторождение расположено на границе Парабельского и Каргасокского районов Томской области вблизи более крупного Лугинецкого нефтегазоконденсатного месторождения.

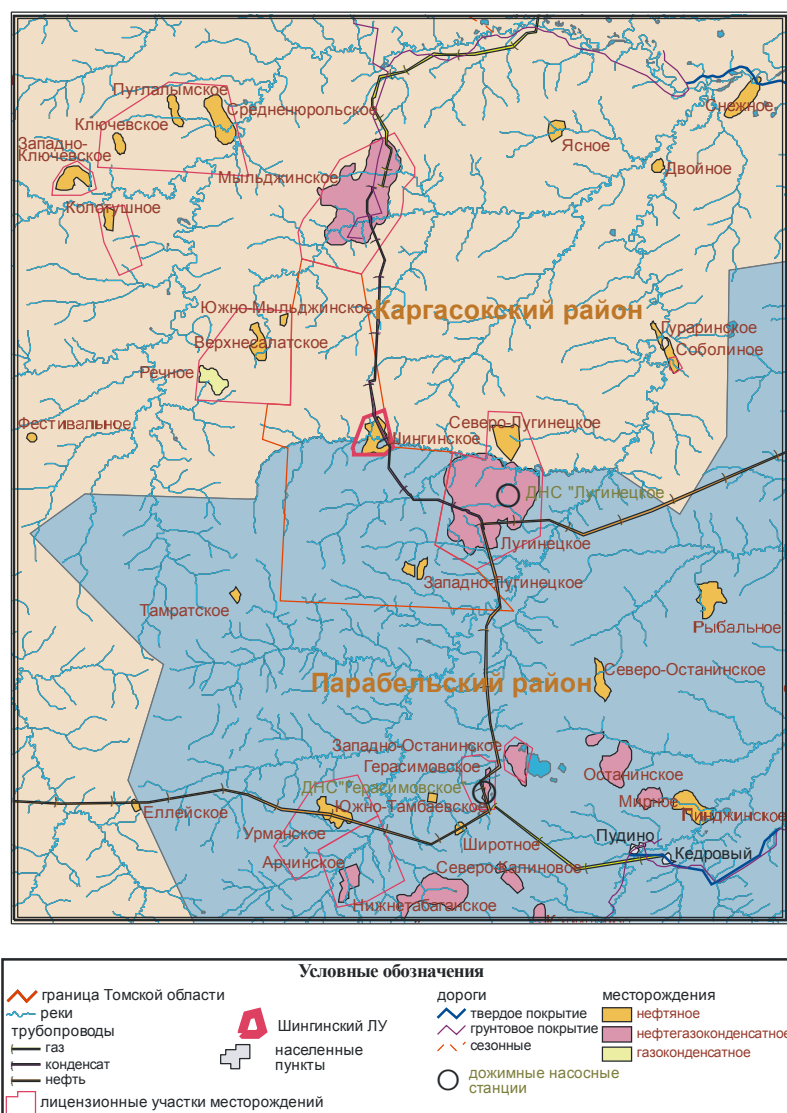


Рисунок 1 – Обзорная карта территории

Пос. Каргасок и Парабель являются районными центрами и расположены в 170 и 150 км от месторождения соответственно. Что касается

населенных пунктов, то ближайшими из них являются город Кедровый и поселок Пудино, которые располагаются в 100 км к юго-востоку.

Месторождение было открыто в 1971 году поисковой скважиной 295, в которой получили промышленный приток нефти из верхнеюрских отложений васюганской свиты.

После дальнейшего бурения разведочных скважин и оценки перспективности разработки месторождения в 1991 году была введена в эксплуатацию разведочная скважина 301 в рамках пробной эксплуатации продуктивной залежи пласта Ю₁¹⁻².

Что касается лицензионных обязательств по разработки месторождения, оно разрабатывается в рамках двух лицензионных участков, а северо-восточная часть месторождения находится в нераспределенном фонде. Лицензии были выданы Министерством природных ресурсов РФ при поддержке Администрации Томской области. На данный момент лицензия на добычу углеводородов была выдана до 2030 года, компанией – оператором является ООО «Газпромнефть-Восток». Основной лицензированной залежью разработки является залежь, расположенная в пласте Ю₁¹⁻² верхневасюганской подсвиты.

Месторождение является удаленным от доступных путей связи. На территории месторождения отсутствуют шоссейные дороги, а ближайшая железнодорожная станция расположена в городе Томск в 440 км от месторождения. Для транспортировки грузов до территории месторождения используют авиатранспорт, а в период навигации по водным артериям. Также, зимой транспортировка осуществляется по зимнику, который связывает город Кедровый и областной центр город Томск.

Ввиду удаленности месторождения от крупных населенных поселков, местное население представлено слабо и оно в основном занимается заготовкой леса, разведением животного скота, промысловой охотой, а в виду бурного введения в разработку Шингинского месторождения с последнее десятилетие – работой на нефтяных промыслах.

Климат на территории месторождения континентальный, который характеризуется продолжительной и суровой зимой с коротким жарким летом. Наиболее жарким месяцем считается июнь со среднемесячной температурой $+23,9^{\circ}\text{C}$, в то время как средняя температура наиболее холодного месяца составляет $-25,7^{\circ}\text{C}$. Абсолютный минимум температуры был зафиксирован на отметке -53°C , абсолютный максимум – на отметке $+36^{\circ}\text{C}$. Территория месторождения имеет равнинный характер, ввиду чего происходит быстрая смена циклонов и антициклонов, что приводит к резкой смене температур.

Увлажненность почвы и количество выпадения осадков считается умеренным. Наибольшая часть осадков выпадает с мая по октябрь, большая часть из которых – в летнее время. Зимний сезон характеризуется относительной сухостью. Среднее количество осадков составляет 492 мм в год.

Что касается грозových явлений, которые встречаются на территории достаточно часто в летнее время, то большая их часть наблюдается в июле. По всей территории за месяц суммарно наблюдается продолжительность гроз до 27 часов.

Снежный покров появляется в октябре и сходит в начале мая. Средняя высота снежного покрова в первый месяц составляет 3 см, с ноября по январь – 17-37 см, в феврале-марте – 43 см. Среднее число дней со снежным покровом составляет 176.

Преобладающее направление ветров юго-западное и юго-восточное в зимний период, в летний период – северо-западное и северо-восточное. Средняя скорость ветра на территории месторождения на высоте 16 метров составляет 3.8 метров в секунду.

В орогидрографическом отношении площадь может быть отнесена к левобережью реки Чижалка и правому притоку реки Васюган. Район может считаться плато со слабым проявлением холмистости, которое расчленено речными долинами с пойменными террасами. Абсолютные отметки от уровня моря находятся в пределах 83 – 121 м. Река Чижалка судоходна для маломерных судов и является транспортной артерией, по которой происходит доставка

грузов на территорию месторождения, расположена в 40км. На территории самого месторождения протекает река Екыльчак шириной до 15 метров, имеющая многочисленные притоки. Река является достаточно глубокой, и проезд через нее возможен только после строительства переправ. Ледостав начинается в конце октября – начале ноября, а сходит лед в конце апреля – начале мая вместе со сходом снежного покрова. Толщина льда в зимнее время достигает 40см. Территория также является заболоченной. Болотистость представлена многочисленными «окнами», которые являются серьезными препятствиями для продвижения по территории месторождения.

Леса на территории относятся к смешанному типу: сосна, кедр, ель, береза, осина. Высота деревьев может достигать 25 метров при толщине стволов до 50см. Подлесок состоит из кустарников высотой до 5 метров. В настоящее время леса находятся на стадии восстановления после затяжных и частых пожаров в начале 20 века.

1.2 Геолого-геофизическая изученность

Исследование геологического строения участка началось еще в 1948 году. В рамках исследований были проведены аэромагнитная съемка, геологическая съемка, гравиметрическая съемка и сейсморазведка методов отражающих волн. Также было проведено бурение опорных скважин для изучения геологического разреза территории. В результате работ был выявлен ряд структур платформенного чехла первого порядка и составлена тектоническая карта. С 1962 года в районе началось поисковое бурение, которое помогло открыть ряд крупных месторождений. Поднятие целевого месторождения было выявлено также по сейсморазведочным работам и подготовлено под бурение по основному отражающему горизонту (подошва баженовской свиты). В результате дальнейшего поискового бурения и было открыто месторождение в 1971 году [7,9].

На месторождении было испытано несколько перспективных горизонтов в отложениях палеозоя, тюменской, васюганской, тарской и покурской свит. В конечном счете был выявлен один продуктивный горизонт для разработки –

верхневасюганский горизонт Ю1. В 2007 году была проведена 3D сейсморазведка, которая позволила получить точные структурные данные по месторождению [9].

1.3 Тектоническое строение территории

Территория месторождения может быть охарактеризована сложным тектоническим строением. По разрезу выделяют нижнюю часть, которая относится к доюрскому фундаменту, который перекрывается осадочным чехлом. Территория может быть отнесена к герцинскому возрасту развития фундамента. В результате осадконакопления в процессе формирования осадочного чехла были унаследованы консидаментационные структуры. Наиболее полное унаследование происходило в районе формирования позднегерцинской складчатости. Таким образом, говоря о типе унаследованных структур, на месте сформировавшихся синклиналий были унаследованы отрицательные структуры в виде прогибов, на месте антиклинорий – положительные структуры в виде поднятий. Само месторождение расположено в пределах Шингинской мезоседловины в месте соединения Нюрольской и Усть-Тымской впадин, Средневасюганского и Пудинского мегаблоков. Расположенные неподалеку структуры третьего порядка, такие как Мыльджинское и Лугинецкое куполовидные поднятия, осложняют Средневасюганской мегаблок и Пудинское мезоподняtie. Анализ проведенных сейсморазведочных работ показал, что унаследованность структур прослеживается вплоть до середины нижнего мела [7,9,11,15].

Что касается самого месторождения, то оно расположено в области высоких абсолютных отметок, которая представляет собой структурный нос, вытянутый в северо-западном направлении [7,11,15].

На Рисунке 2 можно видеть, что месторождение расположено в положительной структуре четвертого порядка, которая в свою очередь осложняет структуру второго порядка, представленную Шингинской мезоседловиной [7,11,15].

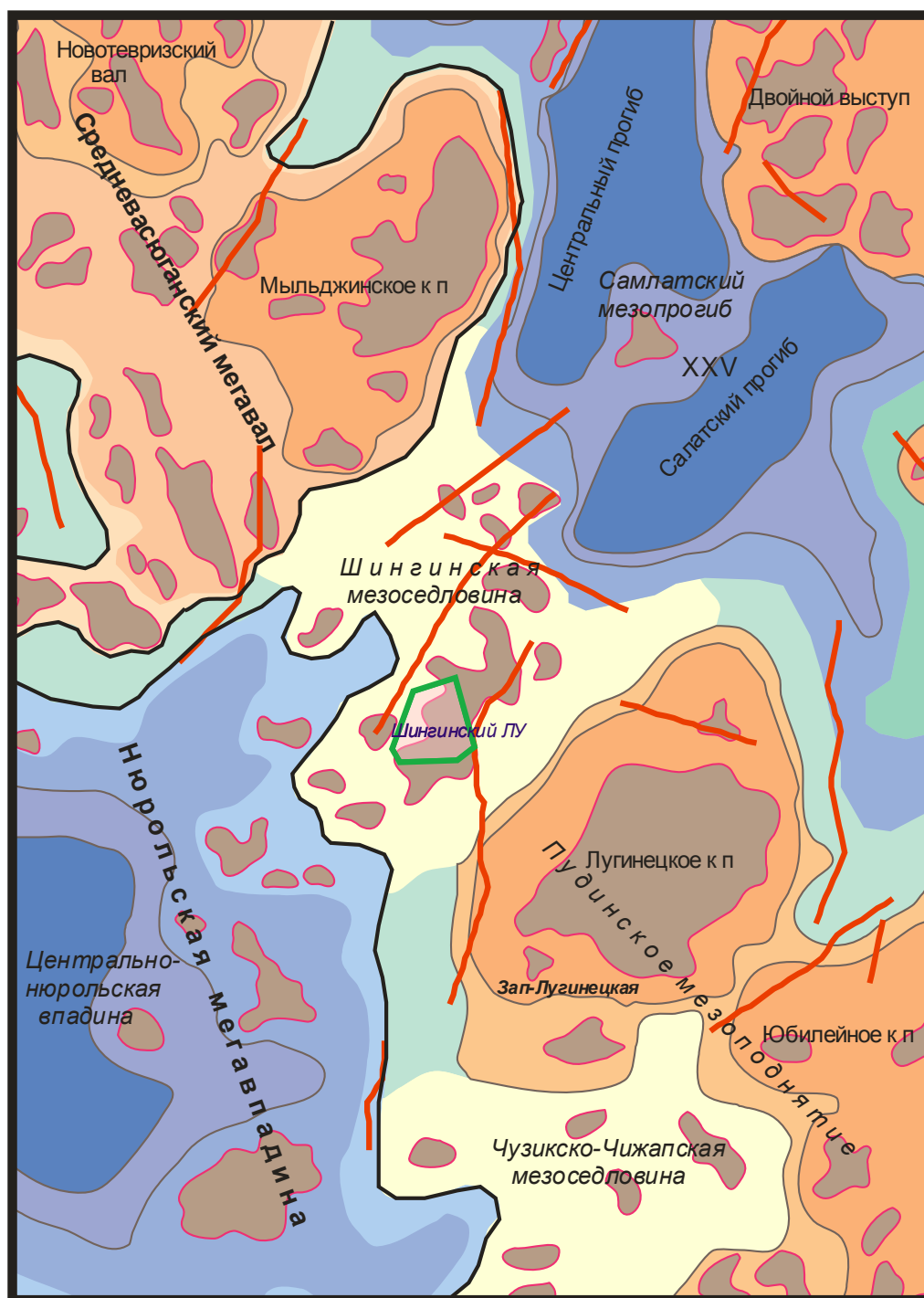


Рисунок 2 — Тектоническая карта юрского структурного яруса

1.4 Стратиграфия месторождения

Как было описано ранее, стратиграфический разрез включает в себя доюрский фундамент и перекрывающий его платформенный осадочный чехол [9,11,15].

1. Палеозойская эратема (PZ)

Нижняя часть палеозойского фундамента сложена кристаллическими сланцами, которые перекрываются карбонатными и глинисто-карбонатными метаморфизованными отложениями девона и нижнего карбона. Отложения представлены светло-серыми глинистыми породами с ярко выраженной трещиноватостью, которая сформировалась в результате метаморфизма первичных отложений. Эти отложения перекрываются вулканогенно-осадочными породами в верхней части фундамента, отнесенных к триасовому периоду осадконакопления, который, как известно, характеризуется ярким проявлением вулканизма. Мощности триасовой системы в рамках отложений колеблются от нуля до десятков-сотен метров и возрастают в погруженных частях. Ввиду малой мощности и локального проявления триасовых отложений их принято относить к фундаменту, а фундамент условно называть сложенным палеозойскими отложениями. Также скважинами были вскрыты толщи пород, и далее был изучен керн, выделены фельзитовые порфиры зеленоватого цвета, которые являются метаморфизованными, выветрелыми.

2. Мезозойская эратема (MZ)

Мезозойско-кайнозойский осадочный чехол в районе месторождения имеет общую мощность 2800-2900 метров. Он несогласно перекрывает фундамент с угловым несогласием, а именно, перекрывает верхнюю часть фундамента, сложенную денудированными и выветрелыми доюрскими породами. На границе фундамент – осадочный чехол прослеживается стратиграфическое несогласие с временным разрывом выделенных осадков. Таким образом, практически полностью отсутствуют отложения пермской системы и в большей степени триасовой, что связано с эрозией осадков с момент поднятия территории. В осадочном чехле выделяются отложения

нижней и верхней юры, которые перекрываются мощными отложениями нижнего и верхнего мела. Выше залегают осадки, сформировавшиеся в палеогеновый, неогеновый и четвертичные периоды.

2.1. Юрская система (J)

В юрской системе выделяются тюменская, васюганская свиты. Также выделяют георгиевскую свиту, представленную в большей степени маломощной барабинской пачкой, которую перекрывает баженовская свита.

2.1.1. Тюменская свита ($J_{2a} - J_{2bt}$)

Тюменская свита представлена только верхней подсвитой, мощность которой составляет 152-171 метров. Подсвита сложена песчаными пластами $Ю_2 - Ю_6$, которые были сформированы в субаквальной обстановке осадконакопления. Песчаные пласты переслаиваются с отложениями аргиллитов, алевролитов, углей и углистых аргиллитов. В разрезе свиты насчитывается до 10 угольных пластов мощностью 1 – 4 метра. Ввиду высокой глинистости песчаных пластов они характеризованы плохими коллекторскими свойствами.

2.1.2. Васюганская свита ($J_{2bt} - J_{3o}$)

По своему строению васюганская свита делится на две ярко выраженных части. Нижняя часть является **нижневасюганской подсвитой**, преимущественно сложенной аргиллитами и алевролитами, не включает в своем строении коллекторов. Песчаные пласты сильно глинизированы и маломощны, поэтому их относят к неколлекторам. В нижней части (перекрывающей тюменскую свиту) выделяют песчаный пласт $Ю_2^0$ (пахомовская пачка), который является базальным пластом позднеюрской трансгрессии.

Верхневасюганская подсвита сложены песчаными и глинисто-алевролитовыми отложениями. Подсвита является нефтегазоносным горизонтом $Ю_1$, с которым связано большинство месторождений Томской области. Подсвита включает четыре песчаных пласта $Ю_1^1 - Ю_1^4$. Песчаники

серого цвета, мелкозернистые, массивные, часто с прослоями аргиллитов и алевролитов.

2.1.3. Георгиевская свита (J_{3o} - J_{3tt})

Свита присутствует локально на территории месторождения, сложена морскими глинами. В некоторых частях территории выделяют барабинскую пачку, сложенную зеленоватыми глауконитовыми алевролитами, которые ярко выделяются по каротажам. Мощность свиты достигает 3 метров в местах проявления.

2.1.4. Баженовская свита (J_{3tt} – K_1b)

Свита сложена глубокоководными морскими битуминозными аргиллитами серого, вплоть до черного цветов, иногда с коричневым оттенком. Встречаются остатки белемнитов и аммонитов, пелеципод. Свита является нефтематеринской породой для большинства месторождений Западно – Сибирской нефтегазоносной провинции. Более того, свита является хорошим отражающим и реперным горизонтом, и достаточно уверенно прослеживается на сейсмических разрезах. Мощность свиты в пределах месторождения достигает 20 метров.

2.2. Меловая система (K)

Меловая система занимает большую мощность осадочного чехла. Мощность системы достигает 2200 метров и представлена система всеми ярусами.

2.2.1. Нижнемеловой отдел (K_1)

В пределах месторождения нижнемеловые отложения состоят из куломзинской, тарской и киялинской свит. Также выделяют покурскую свиту в верхней части разреза.

2.2.1.1. Куломзинская свита (K_1b - K_1v)

В нижней части свита сложена морскими темно-серыми аргиллитами (подачимовская пачка), в верхней части с тонкими пропластками алевролитов (ачимовская пачка), в составе которой выделяются ачимовские песчаные пласты B_{16} – B_{20} . В верхней части свиты встречается песчаный пласт,

формировавшийся в шельфовых условиях B_{10} . Мощность свиты достигает 270 метров.

2.2.1.2. Тарская свита (K_{1v})

Свита сложена в большей степени песчаными отложениями, формировавшихся в прибрежной обстановке осадконакопления, которые включают пласты $B_7 - B_{10}$. Свита характеризуется слабой расчлененностью, а пласты слабо различимы и слагают практически единую песчаную толщу мощностью до 70 метров.

2.2.1.3. Киялинская свита ($K_{1v} - K_{1a}$)

Свита представлена континентальными отложениями. Представлена переслаиванием глинистых пестроцветных пород от красного до зеленого цветов. Встречаются прослой известковистых песчаников и алевролитов. Выделяются маломощные песчаные пласты $B_0 - B_6$ и пласты $A_1 - A_{12}$ мощностью до 6 метров. Песчаные пласты характеризуются линзовидным строением и не являются протяженными по площади. Мощность свиты достигает 600 метров.

2.2.1.4. Покурская свита ($K_{1a} - K_{2s}$)

Свита подразделяется на три подсвиты: нижняя, средняя и верхняя. Нижняя подсвита представлена континентальными отложениями в виде песчаников и алевролитов с прослойками аргиллитов. Выделяют песчаные пласты $ПК_{14} - ПК_{18}$. Средняя подсвита сложена мощными пачками аргиллитов, чередующихся с песчаниками и алевролитов. Отложения формировались в прибрежной обстановке осадконакопления. Включает пласты $ПК_7 - ПК_{12}$. Верхняя подсвита сложена уплотненными песчаниками и алевролитами, выделяют прослой конгломератов. Песчаные пласты – $ПК_1 - ПК_6$. В среднем мощность песчаных пластов покурской свиты составляет 10-15 метров и иногда достигает 30 метров. Общая мощность свиты достигает 900 метров.

2.2.2. Верхнемеловой отдел (K_2)

Верхнемеловой отдел сложен преимущественно отложениями, формировавшимися в морских условиях. Мощность достигает 400 метров.

Отдел состоит из кузнецовской, ипатовской, славгородской и ганькинской свит.

2.2.2.1. Кузнецовская свита (K_2t)

Свита представлена преимущественно глинистыми породами серого цвета с включениями слюдистого вещества и пирита, с растительными остатками. В верхней части свита переходит в алевролитистую породу с содержанием алевролита вплоть до 60%. Толщина свиты доходит до 26 метров.

2.2.2.2. Ипатовская свита ($K_2cn - K_2st$)

Свита сложена песчаными и алевролитовыми породами зеленовато-серого цвета с включениями глауконита, включениями конкреция сидерита и стяжениями фосфоритов. Мощность достигает 160 метров.

2.2.2.3. Славгородская свита ($K_2st - K_2cp$)

Состоит из серых глинистых пород и глауконитовых алевролитов с песчаниками, прослеживаются пиритизированные водоросли, а также включения пирита. Мощность достигает 80 метров.

2.2.2.4. Ганькинская свита ($K_2m - P_1d$)

Состав свиты максимально схож по своему типу со славгородской свитой, за исключением прослоев глинистых опок, которые прослеживаются в ганькинской свите. Мощность достигает 170 метров.

3. Кайнозойская эратема (KZ)

3.1. Палеогеновая система (P)

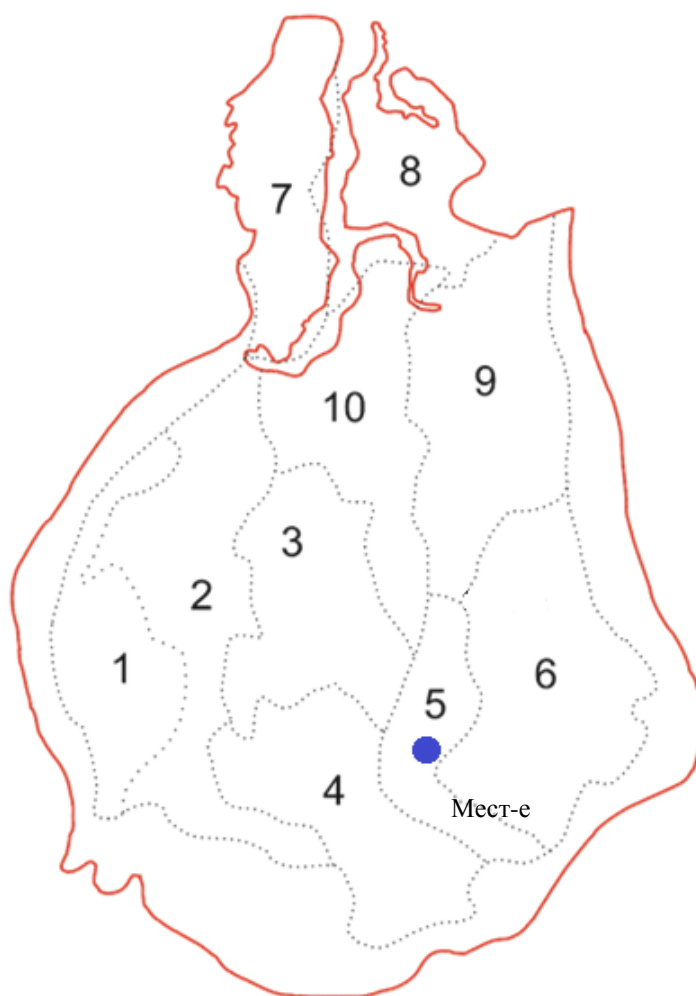
По типу формирования палеогеновая система представлена достаточно широко. Так, встречаются морские отложения, формировавшиеся во время палеоцена и эоцена, так и континентальные, которые были сформированы во время олигоцена. Морские осадки выделяют в талицкой, люлинворской и чеганской свитах. Континентальные отложения относятся к некрасовской серии, формировавшейся во время среднего и позднего олигоцена. Мощность осадков достигает 470 метров.

3.2. Четвертичная система (Q)

Отложения представлены рыхлыми несцементированными осадками, глинами серого и желтоватого цвета, суглинками, алевритами и песками, в верхней части гравием и почвенным слоев. Встречается лигнин и торф. Мощность до 40 метров.

1.5 Нефтегазоносность

Месторождение расположено в Западно – Сибирской нефтегазоносной провинции и относится к Пудинскому нефтегазоносному району Васюганской нефтегазоносной области [9].



Нефтегазоносные области
Западно-Сибирской нефтегазоносной провинции

- | | |
|------------------|----------------------|
| 1 – Приуральская | 6 – Пайдугинская |
| 2 – Фроловская | 7 – Ямальская |
| 3 – Среднеобская | 8 – Гыданская |
| 4 – Каймысовская | 9 – Пур - Тазовская |
| 5 – Васюганская | 10 – Надым - Пурская |

Рисунок 3 – Нефтегазогеологическое районирование Западной - Сибири

Нефтегазоносными горизонтами в районе месторождения являются нефтегазоносные горизонты, расположенные в средней юре, тюменской свите (горизонт Ю2) и в верхней юре, верхневасюганской подсвите (горизонт Ю1). Основным продуктивным горизонтом с промышленными запасами углеводородов, несомненно, верхневасюганский, в котором и находится продуктивный пласт месторождения Ю₁¹ [9]. Также небольшое количество углеводородов было открыто в неокомском нефтегазоносном комплексе. Основное количество залежей на территории связано с антиклинальными структурными ловушками, осложненными литологическими экранами, представленными дизъюнктивами и разрывными нарушениями. В настоящее время на месторождении запасов углеводородом в неокомском и среднеюрском комплексах выявлено не было. Более того, в малой степени был изучен палеозойский нефтегазоносный горизонт, однако в некоторых скважинах, вскрывших, были выявлены следы углеводородов, что позволяет предполагать о возможном интересе к горизонту на месторождении [9]. Тип коллектора палеозойского горизонта является трещинно-каверно-поровый и каверно-карстовый. Породы доюрского комплекса следует связывать с карбонатными постройками рифов, которые в виде пояса рифовой фации прослеживаются на территории интересующей площади. Более того, в связи с перспективностью коры выветривания в качестве коллектора, которая обладает трещинноватостью, полученной в результате денудации и выветривания, следует иметь в виду возможную перспективность данного комплекса помимо рифовых построек. Например, с корой выветривания связана залежь углеводородов Урманского месторождения, а с известняками Северо-Останкинского месторождения – нефтяная залежь. Месторождения расположены в схожих тектонических и геологических условиях [9,11,15].

В виду выделяемых регрессивного и трансгрессивного циклов во время формирования верхнеюрских отложений можно выделить в ее разрезе две пачки, разделяемые угольным пластом: подугольную и надугольную толщи. В

подугольной толщи выделяется пласт Ю₁³⁻⁴. В результате проведенной корреляции было выявлено, что пласт на месторождении включает в себя шесть песчаных линзовидных тел, насыщенных водой мощностью до 30 метров. В надугольной толще выделяют пласты Ю₁¹ и Ю₁². Пласты были сформированы в полифациальных условиях. Пласты разделены небольшой перемычкой, и оба пласта объединяют в один эксплуатационный объект ввиду гидродинамической связанности [9].

Анализ ачимовской свиты показал, что в данной территории не прослеживаются выраженные клиноформы, отражающие сейсмические горизонты имеют субгоризонтальное положение, что говорит о территории находилась в зоне шельфового залегания ачимовской толщи.

Таким образом, объектом разработки месторождения считается верхневасюганская залежь в пласте Ю₁¹. По типу ловушки она относится к структурной литологически изолированной с северо-восточной и западной сторон. Мощность достигает 22 метров. В северо-восточной части находится участок с ухудшенными фильтрационно-емкостными свойствами. Режим работы залежи является упруговодонапорным. Месторождение считается мелком по количеству запасов углеводородов, характеризуется сложным геологическим строением в виду фациальной площадной изменчивости [7,9,11,15].

2 ЛИТЕРАТУРНЫЙ ОБЗОР

В настоящее время нефтяная индустрия перешла из эпохи «легкой нефти» в эпоху, где добыча каждой тонны нефти и кубометра газа – довольно сложная и требующая огромного количества знаний и опыта задача. Если в 20 веке достаточно было найти нефтяную залежь, пробуравив поисковую скважину в купол структуры, выявленной в результате сейсморазведочных работ, а затем составить достаточно простую модель месторождения и проект разработки, который включал бурение вертикальных скважин в рамках регулярной сетки, то на сегодняшний день качество разрабатываемых коллекторов существенно ухудшилось. Теперь мы имеем дело с низкопроницаемыми маломощными пластами с высокой степенью расчлененности и фациальной неоднородностью. Разработка таких объектов требует применения бурения горизонтальных скважин, зарезку боковых стволов, повсеместное проведение ГРП и МГРП. Все это требует от инженера нефтяника достаточной подготовки и умений, что не всегда находится в доступности. Более того, когда мы вступаем в эпоху сложных коллекторов, применение общепринятых подходов при работе с месторождениями становится затруднительным. Особенно это касается карбонатных коллекторов, где применение стандартных подходов интерпретации ГИС просто не работает, а нестандартные подходы не так распространены и специалистов в области карбонатный резервуаров очень мало.

Более того, не секрет, что нефтяная индустрия работает с огромными базами данных, так как компании одновременно владеют большим количеством месторождений с запасами разной сложности, а обрабатывать все эти данные приходится малому кругу специалистов нефтяных и газовых научных центров. Поэтому прогресс привел индустрию к новомодным ветвям развития, таким как «Data Science» и машинное обучение. Сегодня можно с уверенностью сказать, что хороший инженер нефтяник с амбициями должен уметь преодолевать сложные препятствия, а это требует от него знаний не только общих и

геологических, но и углубленных знаний в естественных науках, особенно математики, программирования и машинного обучения, чтобы иметь возможность применять сложные, но действенные способы решения задач.

Зачастую, многие значимые решения, которые принимаются сотрудниками компаний, основаны на опыте специалистов и их «экспертном решении». Не всегда стоит доверять беспрецедентно специалистам, и желательно иметь другие способы, подтверждающие эти решения. Одним из таких способов можно считать алгоритмы машинного обучения, которые при грамотном использовании способны, во-первых, упростить жизнь специалисту, так как ему не приходится больше тратить времени на решение рутинных задач, большинство из них можно оптимизировать с помощью этих алгоритмов, а результаты они дают зачастую более надежные, чем специалисты, так как человеческий фактор здесь отсутствует, во-вторых, существуют задачи, которые специалист просто не может решить, так как возможности человека ограничены. К таким можно отнести, например, анализ и интерпретацию ГИС. Для сложных коллекторов машинное обучение позволяет находить скрытые связи среди набора из 10-20 каротажей, и сравнивать их между сотнями тысяч измерений, прослеживая закономерности. Сделать это человек попросту не может, в конечном счете, в зависимости от способностей интерпретатора, он будет давать ответ на основе двух – трех каротажей, которые, на его взгляд, наиболее хорошо отражают его идею и опыт. Таким образом, мы приходим к неспособности человека работать со сложными объектами в виду его ограниченности способностей. Еще один пример, который наиболее ярко отражает плюсы машинного обучения в нефтегазовой отрасли – это подбор скважин кандидатов для ГРП. Если на месторождении более тысячи скважин, из которых добрая половина – добывающие, а требуется отобрать 2-3 скважины для операции, эта задача становится достаточно сложно решаемой. Особенно, когда таких месторождений для анализа более десяти. Здесь очень хорошо может сработать машинное обучение. Человеку лишь требуется определить наиболее важные параметры, которыми он оперирует при своем качественном

анализе, собрать данные об этих параметрах по всему фонду, научить алгоритм тому, что от него требуется на примере десятка скважин, а затем просто получить конечный ответ по всем скважинам, которые он захочет. Очевидно, сколько времени это может сэкономить, которое специалист может потратить на более важные вещи, например разработку и оптимизацию новых подходов при разработке.

В конечном счете мы можем найти множество примеров, где можно получить эффект от машинного обучения в нефтяной индустрии, начиная с бурения и заканчивая переработкой нефти и транспортировкой ее потребителю. В последнее время была разработана методика восстановления каротажей, которые по какой-то причине не были замерены в скважине, но они нужны для дополнения общей информации, и результаты достаточно обнадеживающие.

В данной работе была разработана методика, новый вариант использования машинного обучения для решения сложных задач. Она включает применение алгоритмов машинного обучения без учителя для выделения продуктивных интервалов в скважинах исключительно по каротажным кривым, минуя процесс интерпретации ГИС, и дальнейший прогноз промышленных характеристик скважин с применением алгоритмов машинного обучения с учителем с целью прогноза накопленной добычи и ранжирования скважин при уплотняющем бурении. Огромный плюс ее заключается в том, что мы минуем процесс построения детальной геологической а затем игидродинамической моделей и длительного расчета скважин на всем фонде. Конечно, модель дает более правильные результаты, так как в ней учтены все факторы, которые могут влиять на добычу, но модель является палкой о двух концах. Если она построена неправильно, то и результаты, какими бы обоснованными они ни были, будут неверными. Поэтому, несмотря на определенные допущения, которые будут перечислены далее в работе, методика может считаться рабочей и действенной. Она может быть использована также и для сугубо геологических задач, а именно, для определения скрытых связей в каротажных кривых, уверенного выделения

нефтенасыщенных коллекторов, коллекторов, различных по фильтрационно-емкостным свойствам, фациального зонирования, как по разрезу, так и по площади. Основной проблемой разработки методики можно назвать проблему кластеризации, решение которой будет детально описано далее. Следует подчеркнуть важность использования алгоритмов машинного обучения без учителя, так как процесс обучения алгоритмов может включать ошибки, допущенные человеком при интерпретации, и алгоритм будет их повторять аналогично человеку, а кластеризация позволяет избавиться от человеческого фактора. В работе далее будет приведено сравнение результатов выделения коллекторов, полученных человеком и машиной, и будет доказано, что зачастую машина делает это лучше и находит то, что упустил человек, например, тонкие пропластки неколектора внутри пласта, которые влияют на фильтрационные характеристики, особенно на вертикальную анизотропию проницаемости.

Поиском решения схожей задачи занимались Джоэл Гевиртц и Адриана Овалле, сотрудники компании Halliburton. Они предложили применять алгоритмы машинного обучения без учителя для фациального деления разреза в скважинах. Основным их мотивом для этого являлось желание снизить возможность ошибок в интерпретации ГИС, которую проводит интерпретатор, и, что самое главное, снизить влияние предубеждений интерпретатора относительно того, как проводить интерпретацию. Зачастую, берясь за новый объект, интерпретатор отражает в результатах интерпретации свой опыт, полученный ранее на других объектах, и занимается поиском известных ему закономерностей в каротажах, что в корне неправильно, так как каждый новый объект неповторим и имеет свои собственные закономерности при интерпретации [17]. Так, фациальное деление может быть основано на форме кривых ПС по методу Муромцева, и результат может быть ошибочным только потому, что человек видел такую форму кривых на другом объекте и там это отвечало определенной электрофации (например, барового тела), поэтому здесь он также выделяет ее, а это, вероятно, не всегда так. Помимо этого, Гевиртц и

Адриана хотели получать результаты интерпретации на основе самих каротажей, а не на основе полученных результатов интерпретации (пористость, проницаемость, водонасыщенность). Все это привело их к разработке методики выделения фаций, вскрытых скважинами, с помощью машинных алгоритмов без учителя [17].

В конечном счете, они разработали методику выделения фаций в скважинах по самим каротажам и описали ее в своих статьях. Например, в своей статье SPE от 2017 года они показали, как методика работает на примере нижнепермских отложений Делаверского бассейна в западном Техасе, блок Wolfcamp A. Этот блок был выбран, так как он представлял наибольший интерес с точки зрения нетрадиционной добычи и геологической сложности разреза. На примере 12 выбранных скважин они применили свою методику [17].

Следует отметить, что в своей работе при подготовке данных ГИС они руководствовались принципом максимальной схожести абсолютных значений каротажей, так как подавали их абсолютные значения алгоритму кластеризации, поэтому они провели детальный анализ подобия каротажей. Такой подход не всегда работает, особенно когда речь идет не о 12 скважинах, а обо всем месторождении. Именно все скважины на месторождении были использованы в данной магистерской диссертации. В разных скважинах замеры ГИС зачастую проводятся разными подрядными организациями на разных приборах, которые могут давать отличные границы абсолютных значений, поэтому их подход в работе с абсолютными значениями не совсем корректен, в настоящей работе проводится нормализация данных, после которой не имеет значения максимальное подобие абсолютных значений каротажных кривых [17].

Для анализа в своей работе они остановились на следующих каротажах: Гамма Каротаж (ГК), Спектральный Гамма Каротаж (СГК) на калии, тории и уране, плотностной каротаж (ГГК-П), Индукционный Каротаж (ИК), Нейтронный Каротаж (НК), калиброванный на пористость, индекс фотоэффекта

(РЕ) и пара Акустических Каротажа на продольных и поперечных волнах (АК) [17].

В работе они предлагают использовать достаточно нестандартный подход: из исходных данных они выделили тренировочный сет, к которому применили алгоритм Агломеративной кластеризации, который входит в группу алгоритмов, основанных на иерархии данных [17].

Выбор количества кластеров на тренировочном сете в их работе был основан на дендрограмме, после анализа которой они остановились на пяти кластерах (Рис. 4). Дендрограмма отражает работу самого алгоритма, его динамику в статике. Принцип работы алгоритмов, основанных на иерархии данных, будет изложен далее в работе более детально. При расчете матрицы расстояний они использовали Манхэттенское расстояние [17].

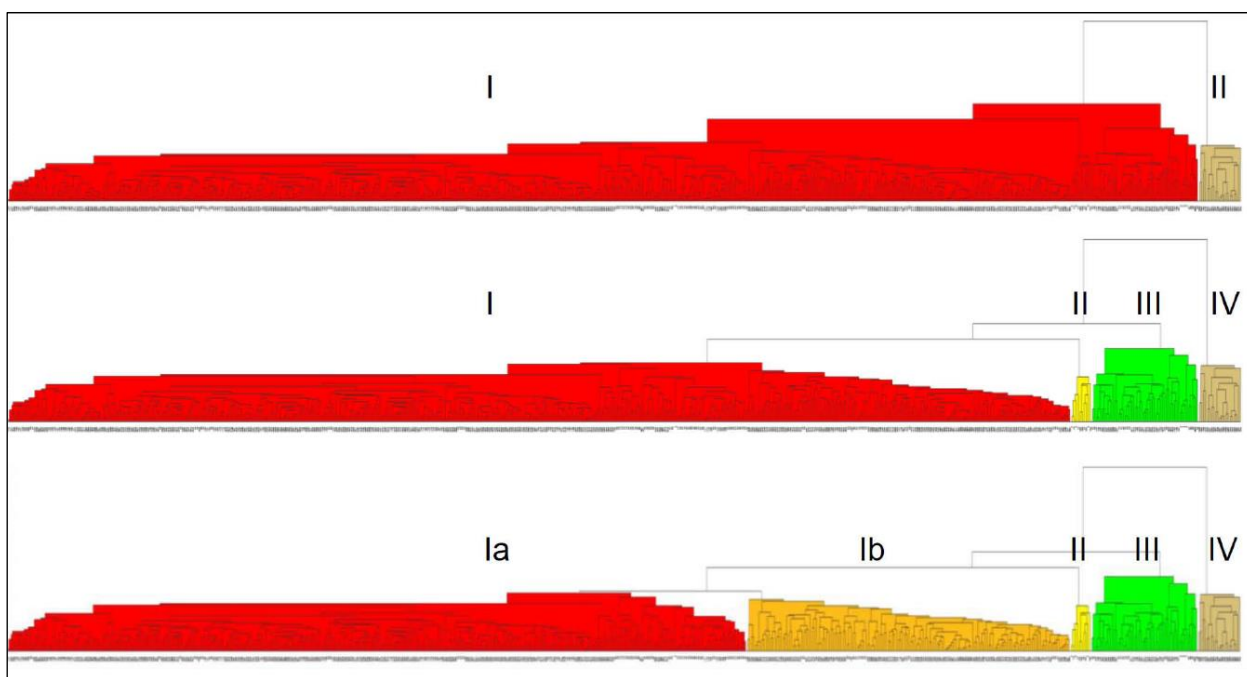


Рисунок 4 – Выбор количества кластеров с помощью дендрограммы

После проведенной кластеризации на тренировочной выборке они применили классификацию оставшихся данных в соответствии с подобием каротажных кривых и их абсолютных значений. Такой подход, вероятно, связан с тем, что на всей выборке результат кластеризации давал плохие результаты, поэтому был сделан такой шаг. Основным недостатком этого подхода является наложение ошибок на каждом шаге. Во-первых, кластеризация геологических

данных не всегда очевидна, так как проводится в многомерном пространстве и не может быть оценена визуально. Следует применять метрики оценки качества кластеризации для обоснования параметров кластеризации помимо визуального анализа дендрограммы, при кластеризации всегда будут аномалии (некоторые наблюдения будут отнесены не в свой кластер, если параметры выбраны необоснованно, например, число кластеров). Во-вторых, при классификации получаем вторую ошибку, так как практически невозможно получить 100% совпадения даже на идеальных учебных данных, а для реальных данных эта точность будет намного ниже. В данной магистерской диссертации предлагается опустить второй шаг, связанный с классификацией, и проводить кластеризацию на всей выборке со скрупулезной оценкой алгоритмов и выбором конечных параметров, которые будут наиболее эффективно работать на всей выборке [17].

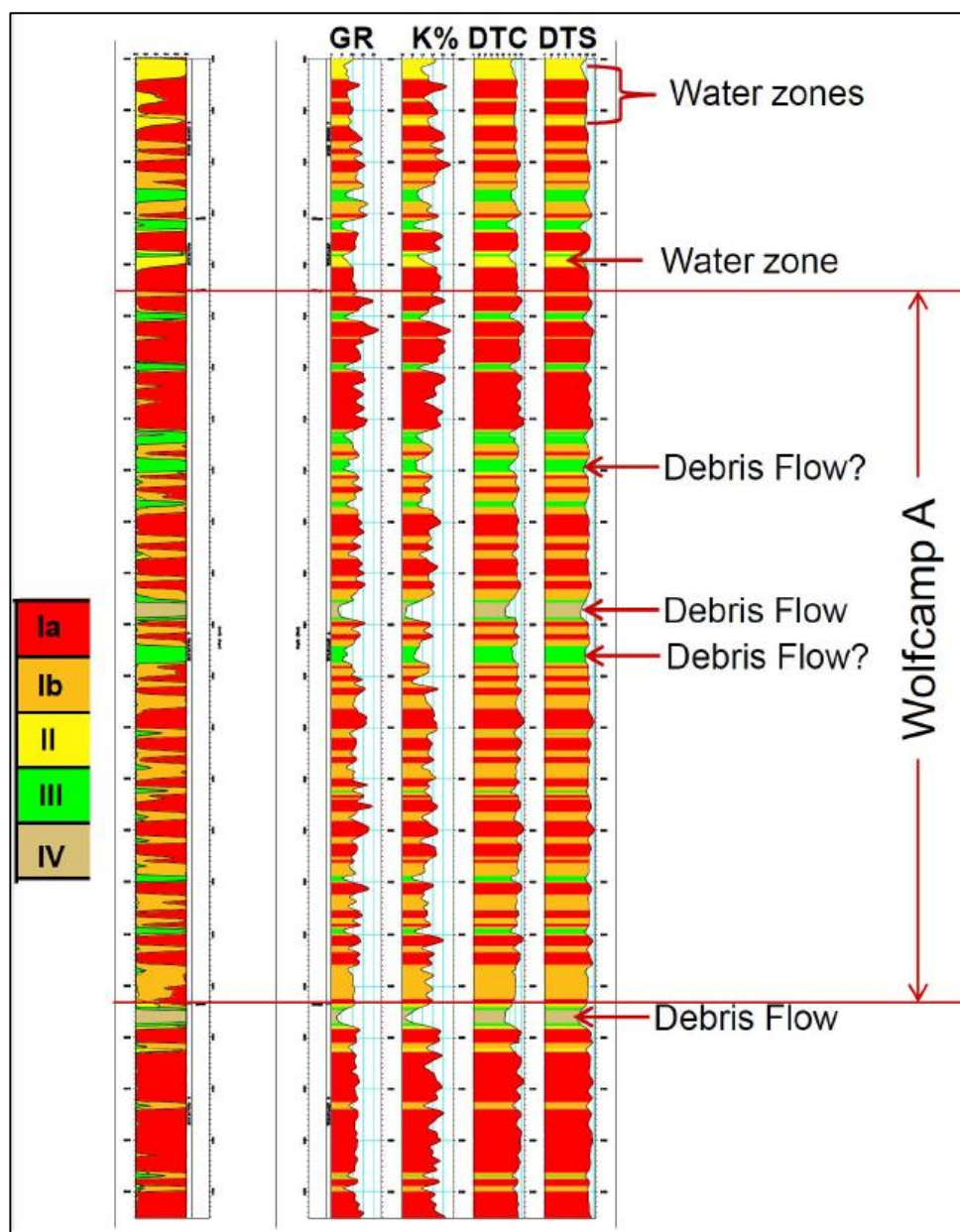


Рисунок 5 – Результат классификации скважины и интерпретация кластеров

В одной из скважин Гевиртц и Овалле имели керновые данные, поэтому у них была возможность обосновать кластеры с геологической точки зрения. Обоснование кластеров и результат кластеризации представлено на Рисунке 5. Как видно, внутри блока Wolfcamp A выделяется переслаивание фаций. Такой результат фациального деления было бы достаточно трудно провести без использования машинного обучения. Можно сделать вывод, что методика требует дальнейшей доработки, которая по возможности была проделана в данной магистерской диссертации.

Помимо кластеризации машинное обучение использовали также для классификации фаций на новых пробуренных скважинах на основе информации выделения фаций в старых скважинах интерпретатором. Брендон Холл в своей работе показывает, как применить этот метод на примере скважин газовых месторождений Хьюгтон и Паномы Северной Америки.

Для классификации он применяет алгоритм Support Vector Machines (SVM), который может быть использован как в качестве классификации, так и для предсказания непрерывных данных. Как классификатор, алгоритм находит наиболее оптимальные гиперплоскости в многомерном пространстве, которые затем используются для классификации новых данных.

На Рисунке 6 показан процесс обучения алгоритма SVM и сравнения обученного алгоритма с колонкой фациального выделения человеком. Следует отметить, что алгоритм смог провести более детализированное вычленение фаций, чем человек, а также можно видеть очевидное сопоставление фаций.

После обучения алгоритма он может использоваться для выделения фаций в любых других скважинах, в которых присутствуют каротажные кривые и делать это достаточно быстро и эффективно.

В данном случае на Рисунке 6 приведены каротаж, которые были использованы: GR – гамма каротаж, ILD – индукционный каротаж, DeltaPHI – разница пористости по нейтронному и плотностному каротажам, PHIND – средняя нейтронно-плотностная пористость, PE – индекс фотоэффекта.

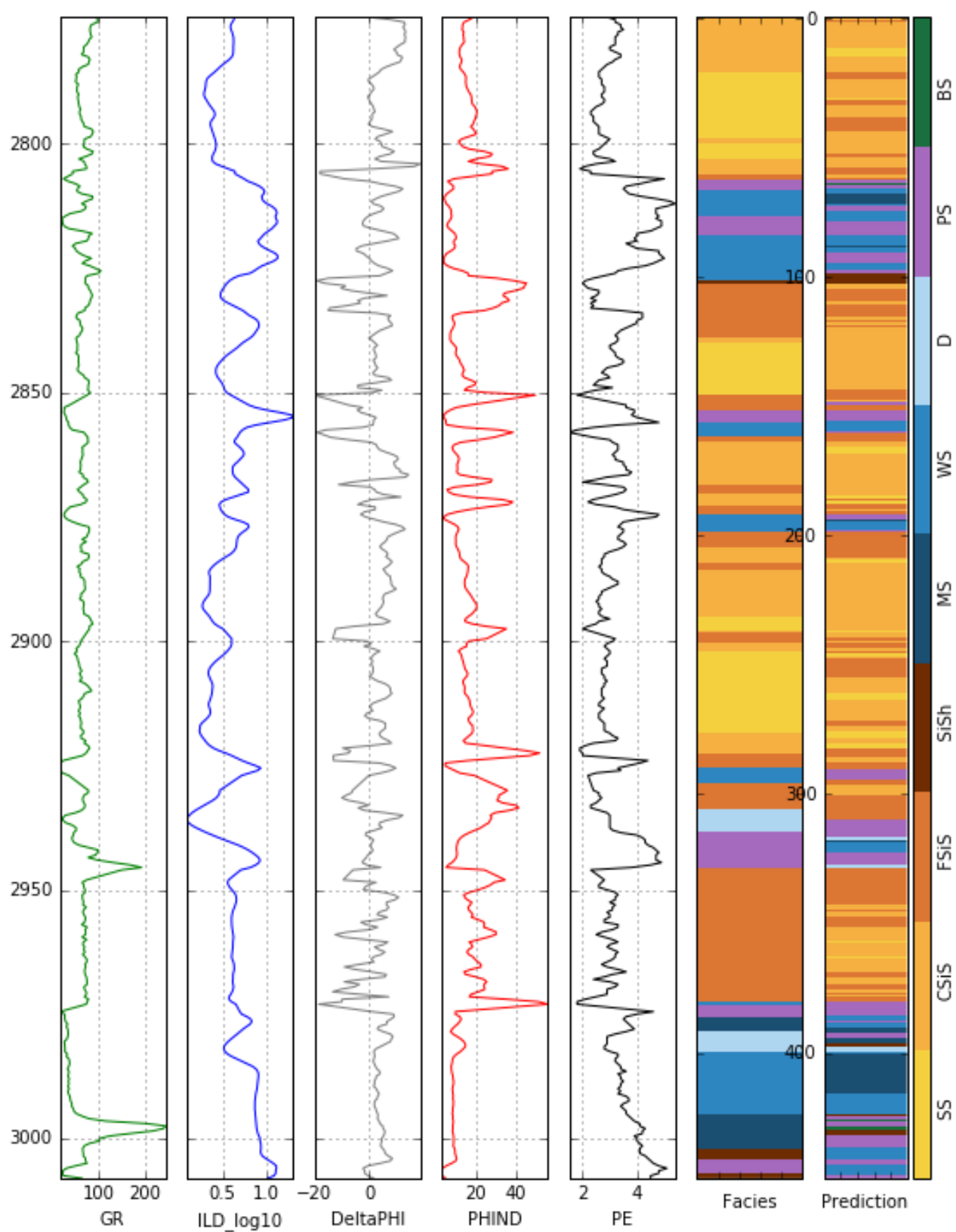


Рисунок 6 – Сравнение выделенных фаций методом классификации с реальными фациями, выделенными человеком

3 ОБЪЕКТ ИССЛЕДОВАНИЯ, ИСХОДНЫЕ ДАННЫЕ И СПОСОБ РЕАЛИЗАЦИИ

3.1 Объект исследования

Основной целью данной работы является разработка и составление комплексной детальной методики по выделению продуктивных интервалов в скважинах с применением алгоритмов машинного обучения без учителя.

На этом этапе планируется усовершенствовать имеющийся подход, разработанный Джоэлом Гевиртц и Адрианой Овалле, а также обосновать его применимость не для конкретного узкого набора скважин со схожими каротажными кривыми, но в более глобальном масштабе в рамках всего месторождения. Эта модернизация позволит применять такой подход для прослеживания фациальной вертикальной и площадной неоднородностей на всем месторождении. Более того, эта методика облегчит процесс подготовки данных для построения геологической модели: позволит интерпретатору обратить внимание на участки, где, возможно, следует провести более детальный анализ при расхождении результатов человеческой и машинной интерпретации, что приведет к построению более надежной геологической модели. Помимо этого, методику можно использовать, когда стандартные подходы фациального зонирования не работают, например, для карбонатных коллекторов доюрских отложений юго-западной части Западно – Сибирской нефтегазоносной провинции, а также при освоении новых месторождений Восточной Сибири.

В работе также предлагается использование результатов кластерного анализа помимо геологических задач для предсказания накопленной добычи нефти и жидкости в рамках составления рейтинга бурения на основе экономических показателей с применением регрессионного анализа.

На сегодняшний день при составлении глобального рейтинга бурения по всем объектам компании используется примитивный расчет, основанный на уравнении Дарси для радиального установившегося притока ввиду

ограниченности доступа к гидродинамической модели и глобального охвата всех объектов. Так как применимость модели становится невозможной, требуется разработка более совершенных подходов для составления рейтинга бурения, чему и посвящена вторая часть данной работы.

3.2 Исходные данные

В качестве исходных данных для выполнения работы использовались каротажные кривые всего фонда скважин месторождения, такие как каротаж Самопроизвольной Поляризации (ПС), Гамма Каротаж (ГК), нейтронные каротаж, разные по принципу датчика, такие как Нейтронный Каротаж на тепловых и надтепловых нейтронах и нейтрон - гамма каротаж. Также использовался Акустический Каротаж на продольных волнах (АК), Боковое Каротажное Зондирование (БКЗ) градиент и потенциал зондами кажущегося сопротивления, Высокочастотное Индукционное Каротажное Изопараметрическое Зондирование (ВИКИЗ). Индукционный Каротаж (ИК), Боковой Каротаж (БК), Микробоковой Каротаж (МБК) микроградиент и микропотенциал зондами.

Также в качестве критерия качества выделения продуктивных интервалов были использованы результаты интерпретации ГИС (РИГИС) по всему фонду совместно с самими каротажными кривыми, в число которых входит выделенная литология, рассчитанная глинистость, пористость, проницаемость и нефтенасыщенность.

Для решения задачи регрессионного анализа были использованы данные по добычи всех скважин месторождения в виде месячных дебитов скважин, времени отработки и накопленных значений флюидов, также информация о пространственной изменчивости пластового давления и забойные давления на скважинах в качестве динамических характеристик.

3.3 Способ реализации

Разработанная методика в полном виде, в каком она представлена в работе, включает следующие основные этапы:

1. Проведение кластерного анализа, в рамках которого требуется обосновать процесс выбора алгоритма кластеризации, а также оптимальных параметров этого алгоритма путем использования математических метрик и визуального сравнения результатов с РИГИС. Этот анализ планируется провести на представительной выборке скважин, наиболее широко отражающих геологическую неоднородность по месторождению и промышленные характеристики скважин, такие как обводенность и запускные дебиты, как по нефти, так и по жидкости.

2. Использование результатов, полученных при кластерном анализе, для решения задачи регрессионного анализа. Требуется обосновать исходные параметры, которые нужно использовать для анализа, выбрать наиболее оптимальный алгоритм регрессии, который позволяет предсказывать запускные дебиты по нефти и жидкости на скважинах в зависимости от геологической ситуации. Также требуется учесть динамические параметры разработки, такие как пластовое и забойное давления, скин-фактор.

3. Обоснование и применение модели темпа падения нефти и жидкости, обоснование фильтрации скважин при таком анализе, формирование представительной выборки для проведения анализа и получение модели падения добычи нефти и жидкости для подсчета накопленной добычи на скважинах.

4. Выбор и обоснование способа интерполяции толщин выделенных в скважинах кластеров в межскважинном пространстве, поиск возможных участков на месторождении для уплотняющего бурения в соответствии с плотностью сетки, отобранными и закаченными флюидами на близлежащих скважинах для составления рейтинга бурения.

5. Предсказание запускных дебитов на проектных скважинах, накопленной добычи нефти и жидкости, расчет экономических показателей на скважинах и ранжирование их в рамках составления рейтинга бурения по месторождению.

4 РАЗРАБОТКА МЕТОДИКИ

4.2 Выбор алгоритмов кластеризации

Во-первых, требуется раскрыть сам термин кластеризация. Кластеризация – это раздел машинного обучения без учителя, где целью является разбиение сета данных на определенное количество групп без предоставленных машине предварительно маркеров и без предварительного обучения. То есть, используются только сами данные, а машина сама находит схожесть между наблюдениями в данных и делит все данные на группы, в которых находятся наблюдения, максимально похожие друг на друга, и максимально непохожие на наблюдения других групп.

Основной проблемой не только при кластерном анализе, а вообще в машинном обучении является выбор самого подходящего алгоритма из большого их множества, который будет наиболее уверенно и точно работать с конкретными данными. Таким образом, огромной ошибкой является необоснованное использование какого-либо алгоритма для сета данных без его обоснования, потому что результаты кластеризации могут быть не такими хорошими, как могли бы быть при использовании другого алгоритма. В результате, мы подошли к проблеме к проблеме выбора алгоритма, когда требуется выбрать максимально разные алгоритмы кластеризации насколько это возможно, применить к ним разные метрики оценки качества кластеризации и выбрать в конечном счете алгоритм, который дает самый точный результат.

Существует четыре основных группы алгоритмов кластеризации:

1. Алгоритмы, основанные на прототипе, то есть общем описывающем центре всего кластера;
2. Иерархические алгоритмы;
3. Алгоритмы, основанные на графах;
4. Плотностные алгоритмы.

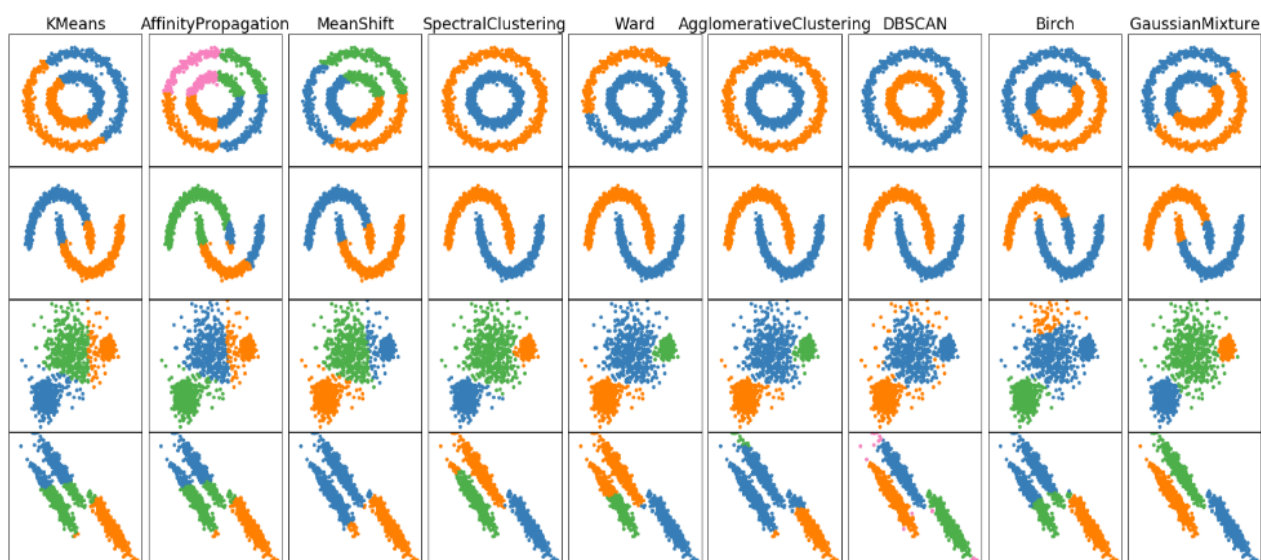


Рисунок 8 – Алгоритмы кластеризации

Из Рисунка 8 очевидно, что разные алгоритмы по-разному работают с разными типами данных. Так как мы практически никогда не знаем, какой тип данных у нас, так как обычно в машинном обучении работают в многомерном пространстве, а данные не могут быть визуализированы, каждый раз при кластеризации мы должны проверять разные алгоритмы на то, как хорошо они работают, и выбирать один наиболее перспективный. В работе выбраны четыре алгоритма, которые характеризуют все четыре группы алгоритмов кластерного анализа, ввиду того что алгоритмы одного типа схожи по принципу действия, и не имеет смысла применять разные алгоритмы одной группы [8,12].

Таблица 1 – Выбранные алгоритмы для проведения кластерного анализа

Группа алгоритмов	Алгоритм
Основанные на прототипе	K-means
Иерархические	Agglomerative clustering
Основанные на графах	Spectral clustering
Плотностные	DBSCAN

После того, как алгоритмы выбраны, необходимо привести описание принципа работы каждого из выбранных алгоритмов.

1. K-means кластеризация

K-means является наиболее популярным и часто применяемым алгоритмом кластеризации. Его применяют повсеместно, так как он является самым простым и, зачастую, эффективным, особенно при работе с многомерными данными, такие, какие участвуют в данной работе [8,12].

Во-первых, следует разъяснить ряд терминов, которые используются в кластерном анализе и машинном обучении в целом. Данные состоят из множества наблюдений и признаков. Каждое наблюдение – это одна точка данных, которая обладает соответствующим значением каждого признака. Например, если бы требовалось применить машинное обучение для классификации группы людей, каждый человек в этом анализе являлся бы наблюдением, которое может быть охарактеризовано разными качествами, то есть признаками, например, вес, рост, возраст, пол, цвет волос и так далее, в зависимости от того, какой результат требуется получить и как нужно поделить людей. Существует также тип данных, называемый маркером. Этот тип данных применяется при классификации и регрессии для обучения. Маркер – это результат, который требуется получить, то есть номер группы людей, либо непрерывное значение. Маркеры используются для обучения модели, чтобы в дальнейшем можно было ее применять на немаркированных данных [8,12].

При кластерном анализе маркеры не используются. Используются только признаки, которые должны присутствовать для каждого наблюдения без исключения. В геологических задачах наблюдениями принято считать точки в скважинах на разных глубинах. Таким образом, для того, чтобы охарактеризовать или найти то или иное наблюдение, мы используем номер скважины и глубину. Признаками являются значения, записанные в каротажных кривых на этой глубине, например, первым признаком является каротаж ПС, вторым – ГК и так далее, в результате чего мы приходим к многомерному пространству, где количество пространств равно количеству каротажей (более 20 в данной работе) [8,12].

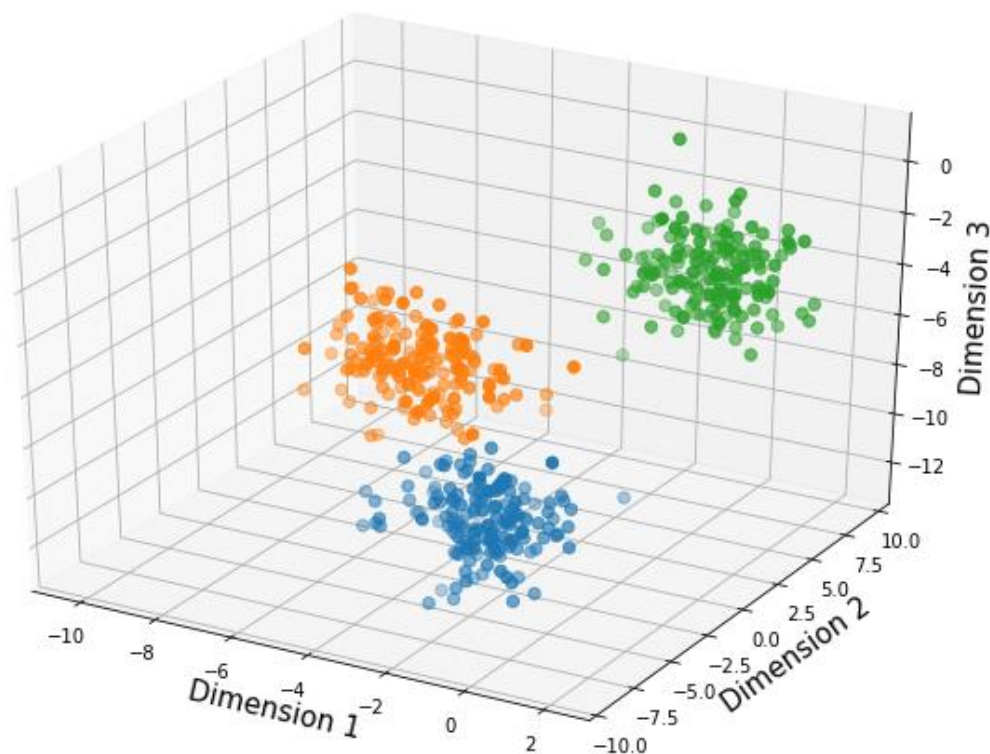


Рисунок 9 – Кластеризация трехмерных данных

Что касается K-means кластеризации, алгоритм относится к типу алгоритмов с прототипом, так как в конце процесса кластеризации каждый кластер остается со своим прототипом, центральной точкой, которая называется центроидом. Также кластер может быть представлен медоидом, если при кластеризации был применен расчет медианного значения вместо среднего. На самом деле, положение центроида задается до начала кластеризации, а затем сдвигается таким образом, чтобы оказаться в центре группы наблюдений. Процедура может быть описана в следующем порядке [8,12]:

1. Задается количество кластеров и фиксируется. Одним из недостатков алгоритма является необходимость задания количества кластеров, обоснование выбора количества кластеров будет приведено далее.

2. Задаются координаты центроидов, число которых соответствует количеству выбранных кластеров. Обычно они задаются случайными координатами, но следует помнить, что если центроиды заданы таким образом,

что слишком далеко находятся от своих конечных оптимальных положений, то процесс кластеризации может существенно затянуться.

3. Задается максимальное количество итераций. Конечно, это не является необходимостью, но если данные являются очень сложными и в них не прослеживаются явно кластеры, то процесс кластеризации может занять очень много времени, а количество итераций может стремиться к бесконечности. Поэтому хорошим тоном считается ограничение количества итераций большим, но конечным значением.

4. Когда запускается первая итерация, происходит расчет расстояний от каждого наблюдения до каждого кластера, после чего каждое наблюдение относится к одному центроиду, который находился наиболее близко к наблюдению. После того, как все наблюдения были разнесены по центроидам, центроид перемещается в центр наблюдений, отнесенных к нему.

5. После того, как центроиды сдвинуты в центр, процедура повторяется в следующей итерации, вновь считаются расстояния до новых положений центроидов, а наблюдения заново относятся к ним. Центроид вновь двигается в центр уже новых отнесенных наблюдений.

6. Итерации прекращаются в тот момент, когда перемещение центроидов относительно их предыдущих положений становится меньше, чем заданное значение погрешности. Обычно значение погрешности задается очень маленьким значением, чтобы получить наиболее точный результат кластеризации.

7. После того, как алгоритм остановлен, каждое наблюдение получает индекс того центроида, к которому они были отнесены в последней итерации. Процесс кластеризации на этом завершен.

Расстояние в алгоритме K-means обычно принято измерять Евклидовым расстоянием, которое считается в многомерном пространстве [8,12]:

$$d(x, \mu) = \sum_{j=1}^m (x_j - \mu_j)^2, \quad (1)$$

где x — наблюдение, μ — центроид, j — номер итерации,
 m — измерение (признак)

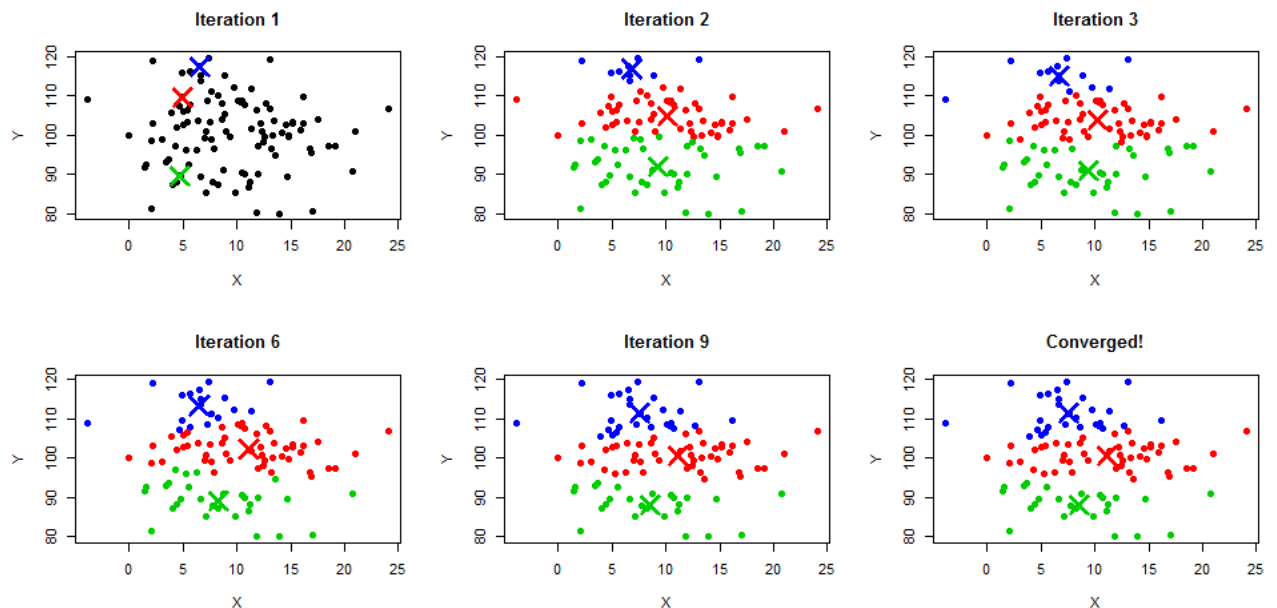


Рисунок 10 — K-means кластеризация в двумерном пространстве с тремя центроидами в девяти итерациях

2. Agglomerative clustering

Агломеративная кластеризация является алгоритмом, основанным на иерархии данных. Основным преимуществом этого алгоритма является то, что можно построить дендрограмму, которая визуальнo отображает процесс кластеризации в не зависимости от количества измерений, из которой можно сделать вывод о необходимом количестве кластеров. Более того, не требуется задавать количество кластеров, можно зафиксировать их количество, но не требуется их задавать, чтобы запустить алгоритм. Процесс кластеризации очень схож с алгоритмом divisive clustering. Отличием является то, что при агломеративной кластеризации процесс начинается с количества кластеров, равному количеству наблюдений, а в процессе кластеризации происходит объединение кластеров, пока не останется один кластер. В алгоритме divisive clustering процесс обратный, процесс начинается с одного кластера и завершается количеством кластеров, равному количеству наблюдений [8,12].

В данной работе будет использован алгоритм агломеративной кластеризации.

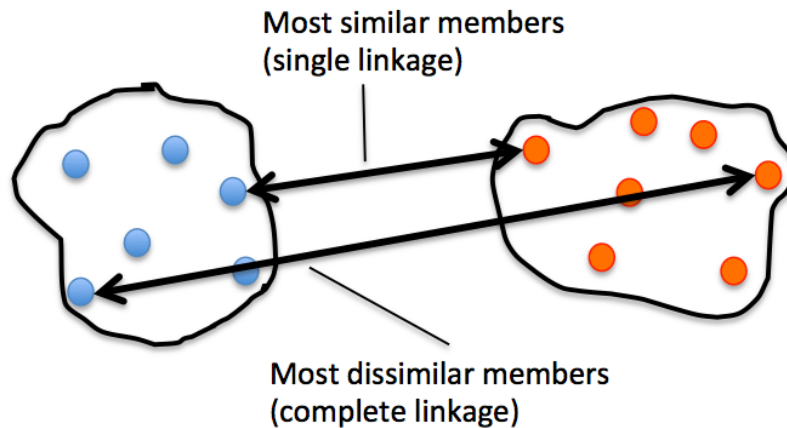
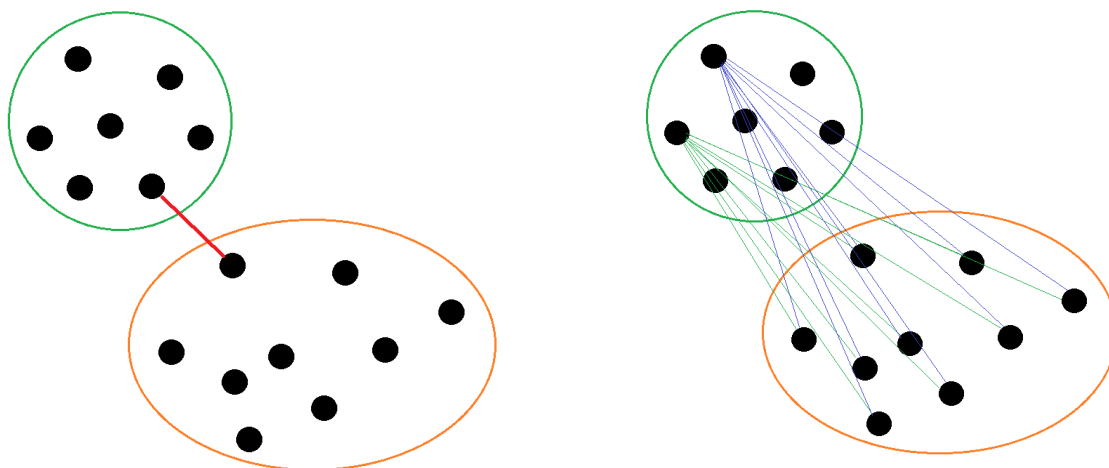


Рисунок 11 – Процесс агломеративной кластеризации

Существует четыре типа агломеративной кластеризации: с простым объединением, с полным объединением, со средним и с типом Ward. В каждом способе считается и строится матрица расстояний, которая включает расстояние между каждыми двумя наблюдениями. В простом объединении сначала объединяются в один кластер кластеры, между которыми наименьшее расстояние, в полном объединении – наиболее отличные кластеры, расположенные дальше друг от друга [8,12].

Что касается объединения средних, то кластеры объединяются в соответствии с наибольшей когезией. Таким образом, в простом объединении мы имеем два кластера на какой-то итерации, которые объединяются в один при условии, что два наблюдения внутри кластеров имеют следующее наименьшее расстояние из матрицы расстояний. Объединение средних более справедливое, так как считается среднее всего кластера, и только потом происходит сравнение расстояний. В типе Ward кластеры объединяются в один при условии наименьшего увеличения суммы квадратов разности внутри кластеров после объединения. Очевидно, что такой подход является самым справедливым и объективным, нежели простое объединение или среднее, например [8,12].



а) Простое объединение

б) Объединение средних

Рисунок 12 – Сравнение типов агломеративной кластеризации

Таким образом, основные шаги кластеризации следующие:

1. Строится матрица расстояний
2. Каждое наблюдение определяется кластером
3. Итерационно кластеры объединяются до тех пор, пока не останется один

3. Density-based Spatial Clustering of Applications with Noise (DBSCAN)

Этот алгоритм основан на обнаружении областей наблюдений с высокой плотностью и объединения этих наблюдений в один кластер. Основным преимуществом является возможность определения шума и исключения его из основных кластеров [8,12].

Процедура кластеризации следующая:

1. Задаются два параметра, такие как радиус исследования (ϵ) и минимальное количество наблюдений (MinPts). Радиусом является расстояние, на котором происходит поиск минимального количества наблюдений;
2. Часть наблюдений относится к **ядрам**, если в заданном радиусе находится большее количество наблюдений, чем минимальное количество.
3. Часть наблюдений относится к **краевым наблюдениям** при условии, что в их радиусе исследования определяется количество наблюдений,

меньше минимального необходимого количества, но они находятся на расстоянии от других ядер, меньшем, чем радиус исследования.

4. Оставшиеся наблюдения, которые не имеют в радиусе исследования необходимое количество других наблюдений и находятся на расстоянии к ближайшему ядру, большем, чем радиус исследования, такие наблюдения относятся к шуму.

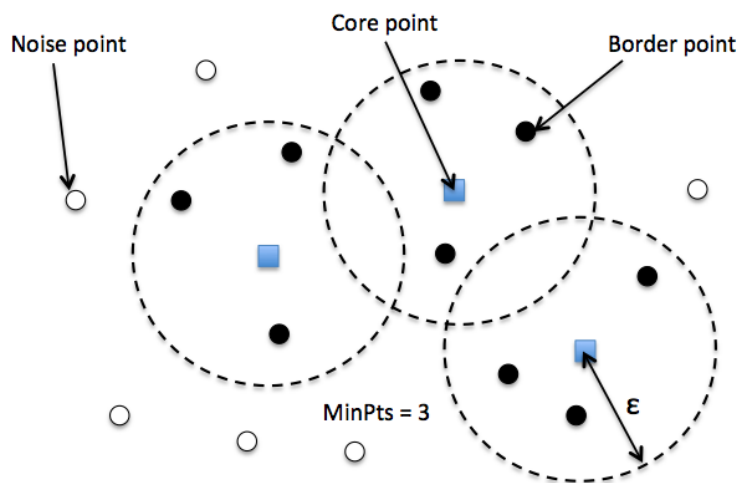


Рисунок 13 – DBSCAN процедура

5. Ядра объединяются в один кластер, если они находятся друг от друга на расстоянии, меньше радиуса исследования, а их краевые точки также причисляются этому кластеру.

6. Шумы относятся в отдельный кластер шумовых наблюдений. Если необходимо, можно определить, какой кластер относится к шуму, и избавиться от наблюдений в данных.

Таким образом, для алгоритма не имеет значения форма кластеров, он все равно сможет провести успешную кластеризацию. Основной проблемой, возникающей при работе с этим алгоритмом, является необходимость задания параметров, таких как радиус исследования и минимальное количество наблюдений. Обычно, это является проблемой, которая требует оптимизации алгоритма для ее решения [8,12].

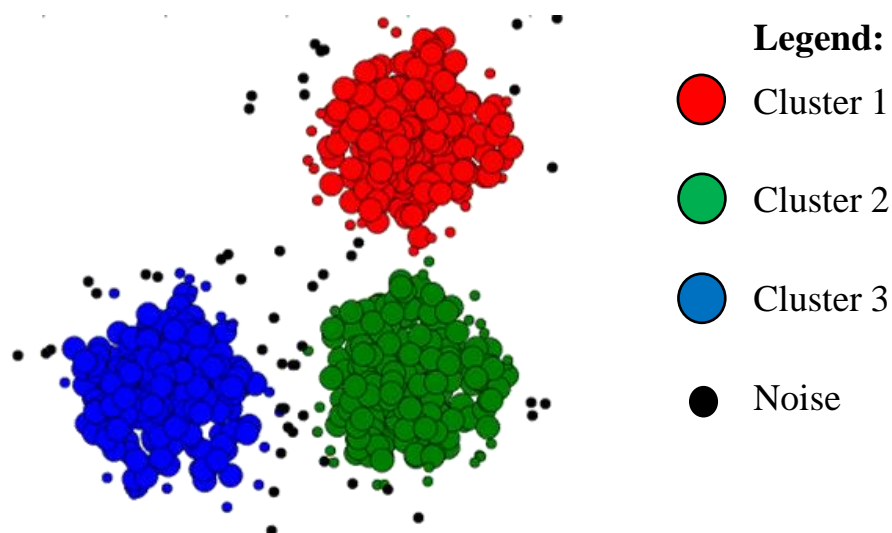


Рисунок 14 – Результат кластеризации алгоритмом DBSCAN

4. Спектральная кластеризация

В процессе кластеризации первым шагом является определение матрицы похожести между каждым двумя наблюдениями [8,12]:

$$A_{ij} = -\left\|x_i - x_j\right\|^2 \quad (2)$$

Такая матрица позволяет описать полный граф, у которого вершины являются точками наблюдений, а ребра соединяют каждую вершину, а весом является степень похожести, занесенная в матрицу похожести вершин. Таким образом, похожесть определяется длиной ребра, а две вершины максимально похожи друг на друга, если расстояние между ними наименьшее [8,12].

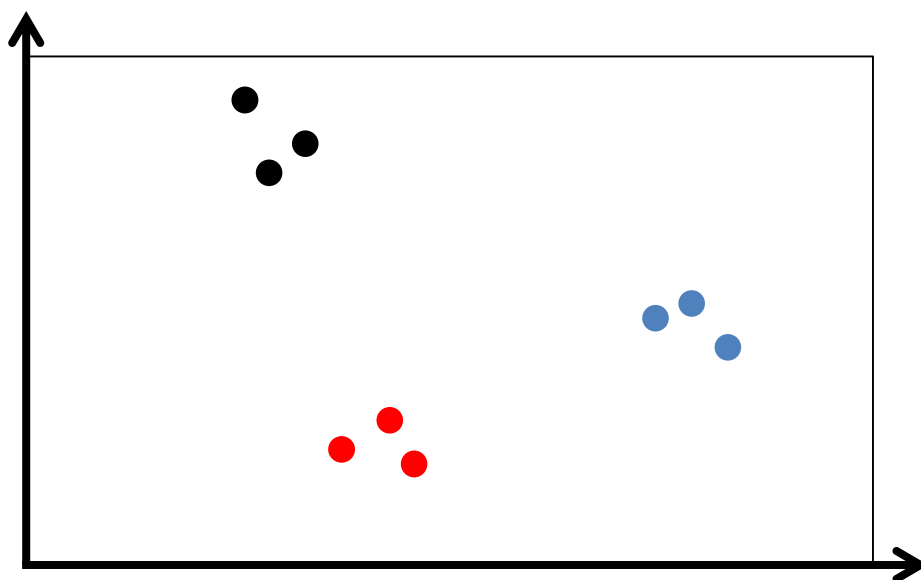


Рисунок 15 – Данные двумерного пространства

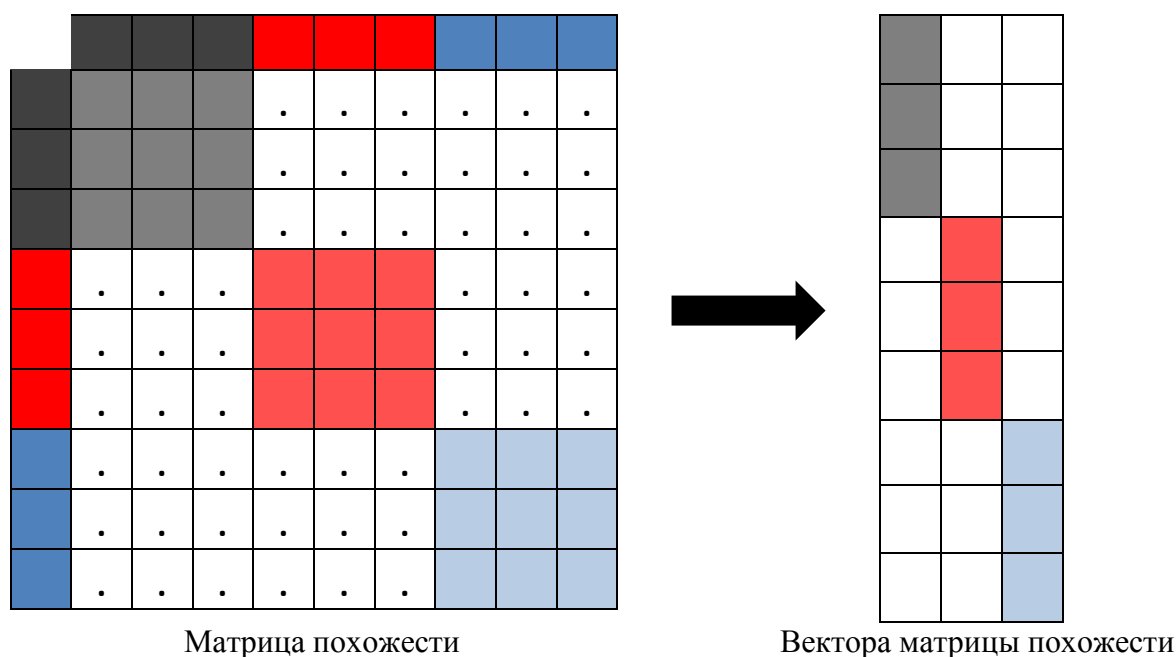


Рисунок 16 – Процесс спектральной кластеризации

После чего происходит пересчет векторов матрицы похожести, которые в дальнейшем используются в ранее описанном алгоритме K-means. Такой подход позволяет более детально описать исходные данные и, в некоторых случаях, это позволяет получить лучшие результаты, нежели алгоритм K-means в чистом виде [8,12].

4.3 Подготовка данных

Очень важным этапом при кластерном анализе является подготовка данных, потому что, если данные являются некондиционными, ни один алгоритм, каким бы совершенным он ни был, не сможет достичь хороших результатов. В литературном обзоре было отмечено, что Джоэл Гевиртц и Адриана Овалле использовали ненормализованные данные. На этом этапе этот недостаток их методики должен быть нивелирован. Таким образом, первым шагом в подготовке данных является загрузка всех многочисленных лас файлов для каждой скважины, участвующей в анализе, и формирование законченной полной таблицы, в которой содержатся данные всех скважин, достаточно хорошо обработанные, которые могут быть использованы в анализе [13,8,12,19].

Вся работа с кластеризацией, загрузкой, обработкой и выгрузкой данных, была проведена с помощью языка программирования Python в среде Jupyter Notebook, так как Jupyter Notebook является очень простым и удобным средством для научного программирования ввиду того, что все получаемые в процессе расчетов результаты тут же могут быть визуализированы [6,13,19].

В работы были использованы данные разного типа, поэтому для работы были использованы различные библиотеки. Библиотека Pandas использовалась для работы с двумерными наборами данных, такими как таблицы. Она очень удобна для загрузки и выгрузки файлов Excel и имеет множество функций для удобной работы с таблицами [6,13,19].

Библиотека Lasio использовалась для загрузки лас файлов, в которых содержатся записанные данные ГИС. Для построения графиков были использованы библиотеки Matplotlib и Seaborn. Для непосредственной работы с алгоритмами машинного обучения были использованы библиотеки Scikit-learn и Scipy. Основной библиотекой является Scikit-learn, в которой содержатся практически все алгоритмы машинного обучения, а также необходимые метрики оценки качества этих алгоритмов, тогда как Scipy использовалась для построения матриц расстояний и дендрограмм при работе с агломеративной кластеризацией [6,13,19].

В первую очередь были загружены каротажные кривые и обрезаны по скважинным отбивкам целевого пласта, так как в работе основной интерес представляет процесс кластеризации именно самого пласта, а не всего разреза.

Следующим шагом являлся поиск общих каротажных кривых среди скважин, так как алгоритмы не могут работать с пропусками в данных. При кластеризации все данные должны быть похожи и иметь одинаковое наполнение [8,12,19].

Несмотря на это, часть скважин (около 10%) была исключена из анализа ввиду того, что на скважинах не были проведены замеры важных каротажных кривых, таких как каротаж ПС, в то время как на 90% остальных скважин он проведен был, и отбросить этот каротаж из анализа было бы некорректно.

Общими каротажными являются: ПС, ГК, НК, БК, ИК, БКЗ (6 градиент зондов), МБК, ВИКИЗ (9 зондов).

Перед тем, как была проведена нормализация данных, аномальные значения в каротажных кривых были исключены, которые вылетают за рамки двух стандартных отклонений, и были заменены значениями, равными среднему плюс минус два стандартных отклонения. Этот шаг очень важен, так как Нейтронный Каротаж показывает аномальные значения на углях, и нормализация не может быть выполнена по аномальному значению. Более того, Индукционный Каротаж и каротаж кажущегося сопротивления могут показывать очень низкие значения на высокоомных или наоборот низкоомных горных породах, таких как мергель, песчаник с включениями пирита. Стандартное отклонение было поскважинно посчитано для каждого каротажа отдельно, и замена аномалий производилась также поскважинно.

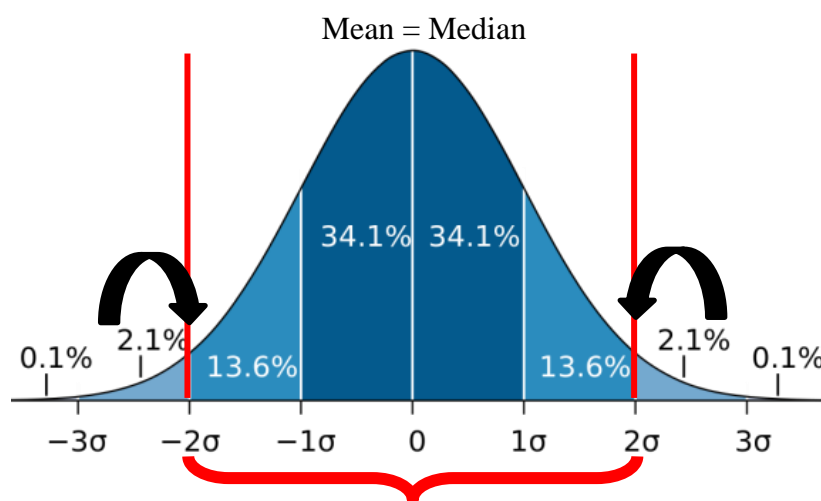


Рисунок 17 – Процесс исключения аномалий из каротажных кривых

После того, как аномалии были исключены из данных, каротажные кривые были поскважинно нормализованы от 0 до 1, чтобы получить все каротажные кривые в одной размерности, так как каротажные кривые в разных скважинах могут иметь схожую форму, но разные абсолютные значения. Из-за этого они будут кластеризованы как разные типы коллектора. В результате не важно, насколько отличаются абсолютные значения, они будут правильно разбиты на группы.

Последний шаг в подготовке данных – сбор нормализованных каротажных данных в одну таблицу, которую можно использовать при кластерном анализе [6,8,12,19].

4.4 Разработка метрики оценка качества кластеризации

Самой сложной частью кластерного анализа является плавильный выбор и ранжирование алгоритмов. Наиболее обоснованным способом является проверка того, как хорошо алгоритм кластеризует данные путем сравнения результатов кластеризации с предварительно размеченными данными. В настоящей работе такой возможности нет, поэтому требуется применение математических метрик, которые оценивают результат кластеризации разных алгоритмов, а затем они сравниваются [20].

Существует несколько метрик, например, для оценки алгоритма K-means используется метод локтя, при котором считается дисперсия данных внутри кластера. Недостатком такой метрики является необходимость в координатах центроидов, которых нет в других алгоритмах. Также существует метод Силуэт, но он часто дает нестабильные результаты и не позволяет достоверно выбрать количество кластеров, или оценить качество кластеризации [8,12,20].

Эта проблема была решена модернизацией метода локтя таким образом, чтобы метрику можно было применить ко всем алгоритмам и сравнить результаты [8,12,20].

Алгоритм применения метрики следующий:

- 1) задается количество кластеров, в первой итерации равное двум;
- 2) проводится кластеризация выбранным алгоритмом;
- 3) строится матрица расстояний для каждого кластера отдельно;
- 4) считается среднее расстояние внутри каждого кластера;
- 5) считается среднее расстояние внутри кластера для итерации;
- 6) в следующей итерации увеличивается количество кластеров;
- 7) строится график зависимости средних расстояний от количества кластеров.

Решение проблемы определяется графически. Оптимальным значением кластеров считается такая точка, после которой среднее расстояние начинает меняться менее интенсивно, то есть дальнейшее увеличение числа кластеров влияет незначительно на точность кластеризации.

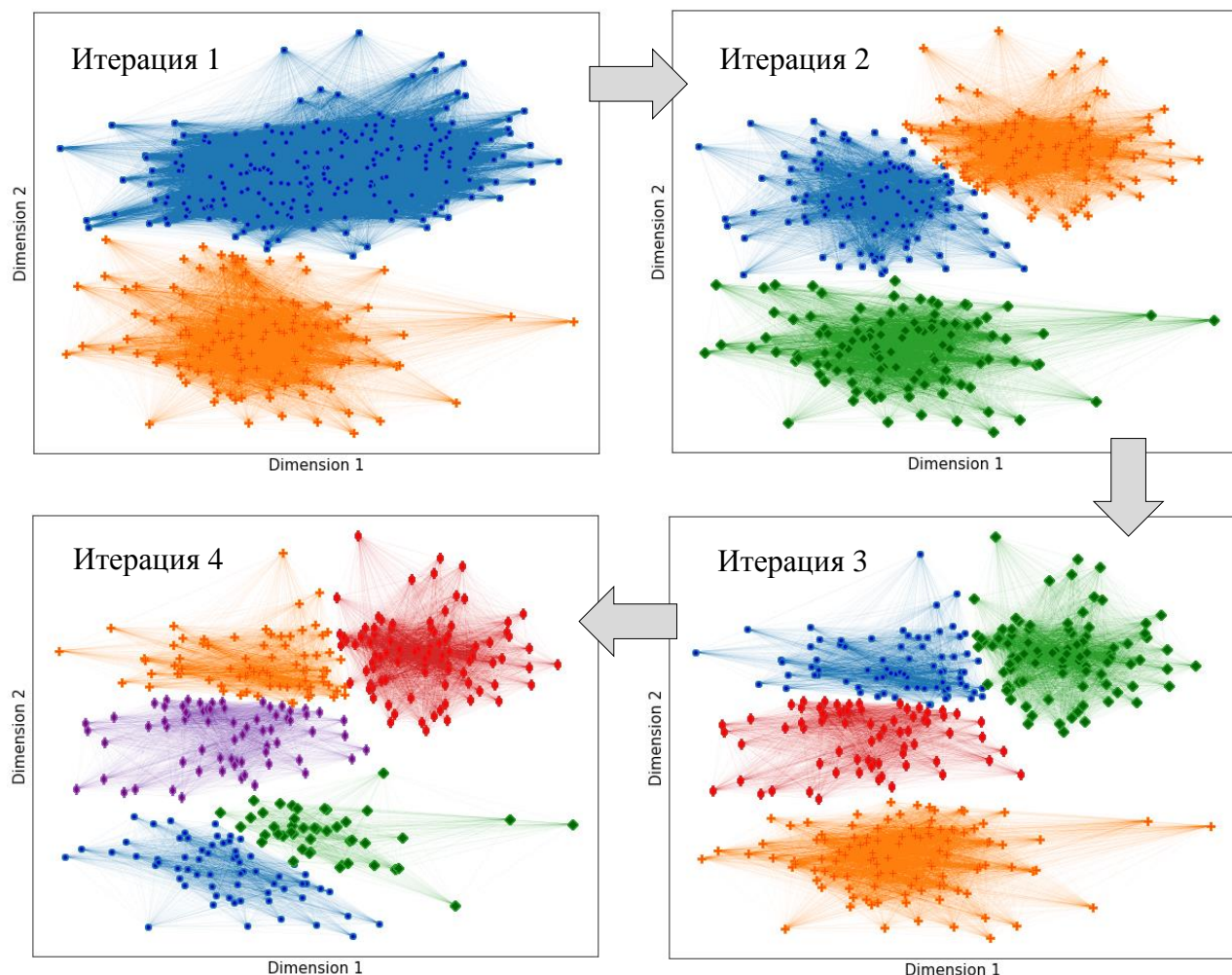


Рисунок 18 – Итеративный процесс определения оптимального количества кластеров

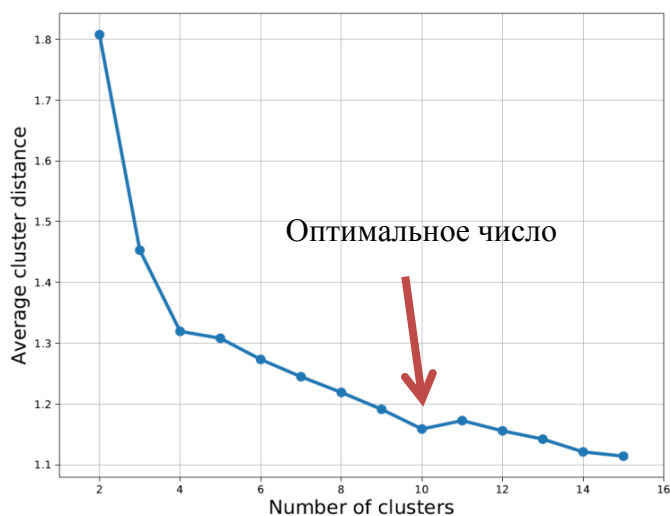


Рисунок 19 – Оценка оптимального количества кластеров

Метод оценки средних расстояний был протестирован на различных двумерных наборах данных, на которых метрика показала достаточно хорошие результаты, поэтому в дальнейшем она была использована как базовая для процесса оптимизации [20].

Выбранные алгоритмы кластеризации были применены для оценки оптимальных параметров с применением метрики.

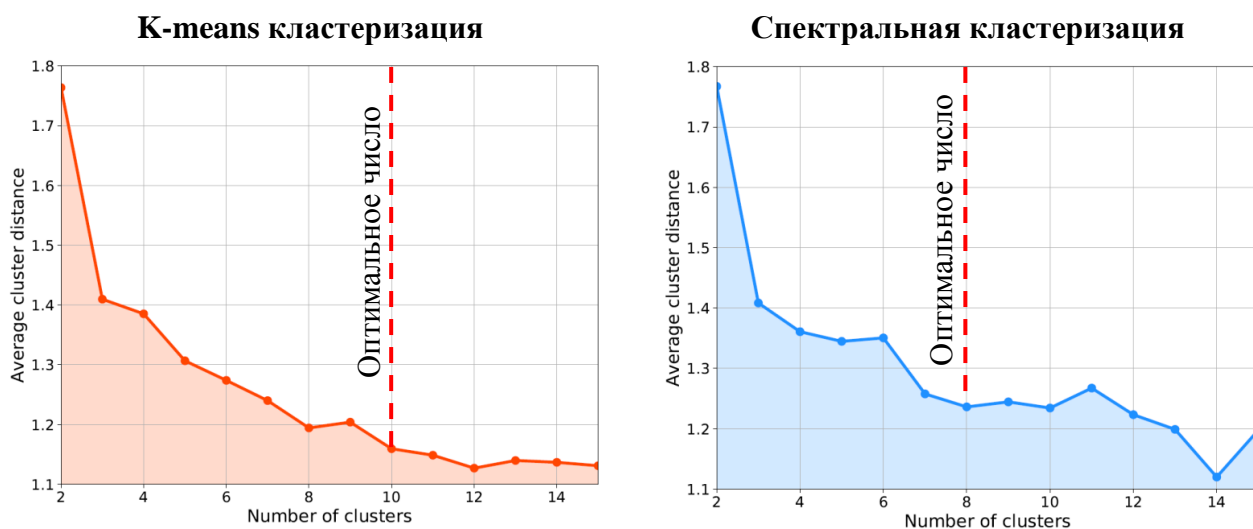


Рисунок 20 – Результат выбора числа кластеров для K-means и спектральной кластеризации

Из Рисунка 20 видно, что оптимальное число кластеров для алгоритма K-means – десять. Количество кластеров, равное восьми тоже является привлекательным, но после оценки кластеризации по восьми кластерам визуально, полученные результаты были недостаточно хороши, поэтому было выбрано число десять.

Что касается спектральной кластеризации, то восемь кластеров являются оптимальным значением и достаточно хорошо показывают себя при визуальной оценке.

Что касается DBSCAN, то требуется оптимизация двух параметров, таких как радиус исследования и минимальное количество наблюдений. Оптимизация была проведена таким образом, что евклидово расстояние было задано в границах 0.1 и 0.9 с шагом 0.1, для каждого из которых было определено оптимальное минимальное количество наблюдений, после чего результаты для разных радиусов исследования были сравнены. Количество кластеров при подборе было ограничено в границах 5 - 15, так как другие

алгоритмы показали недостаточность количества кластеров, меньше пяти, а количество кластеров более 15 просто нецелесообразно.

График оптимизации представлен на Рисунке 21. Оптимальные параметры были выбраны при радиусе исследования 0.2 и минимальном количестве наблюдений, равном десяти (при восьми кластерах). После чего, визуальная оценка кластеризации показала некондиционные результаты. Связано это с тем, что в отдельный кластер заносятся все шумы, а они мешают проводить достоверную математическую оценку. В результате, выбор параметров был проведен вручную визуально, анализируя сходимость результатов кластеризации с РИГИС. Выбранные параметры: радиус исследования равен 0.6 при минимальном количестве наблюдений, равном 20 [8,12,20].

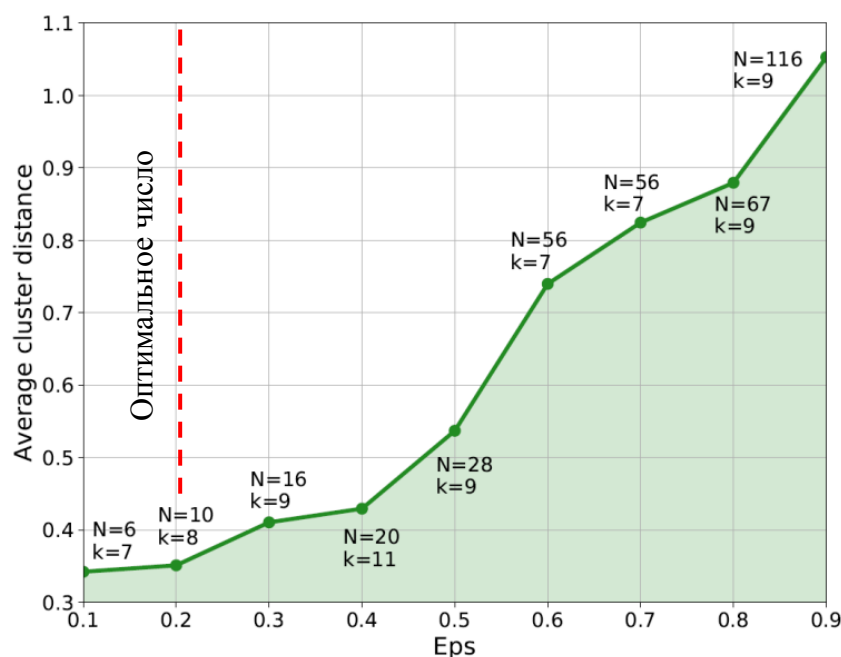


Рисунок 21 — DBSCAN оптимизация

Агломеративная кластеризация позволяет оценить алгоритм не только математически, но и визуально с помощью дендрограммы, которая позволяет оценить динамическую кластеризацию в многомерном пространстве и проследить динамику процесса статически [8,12,20].

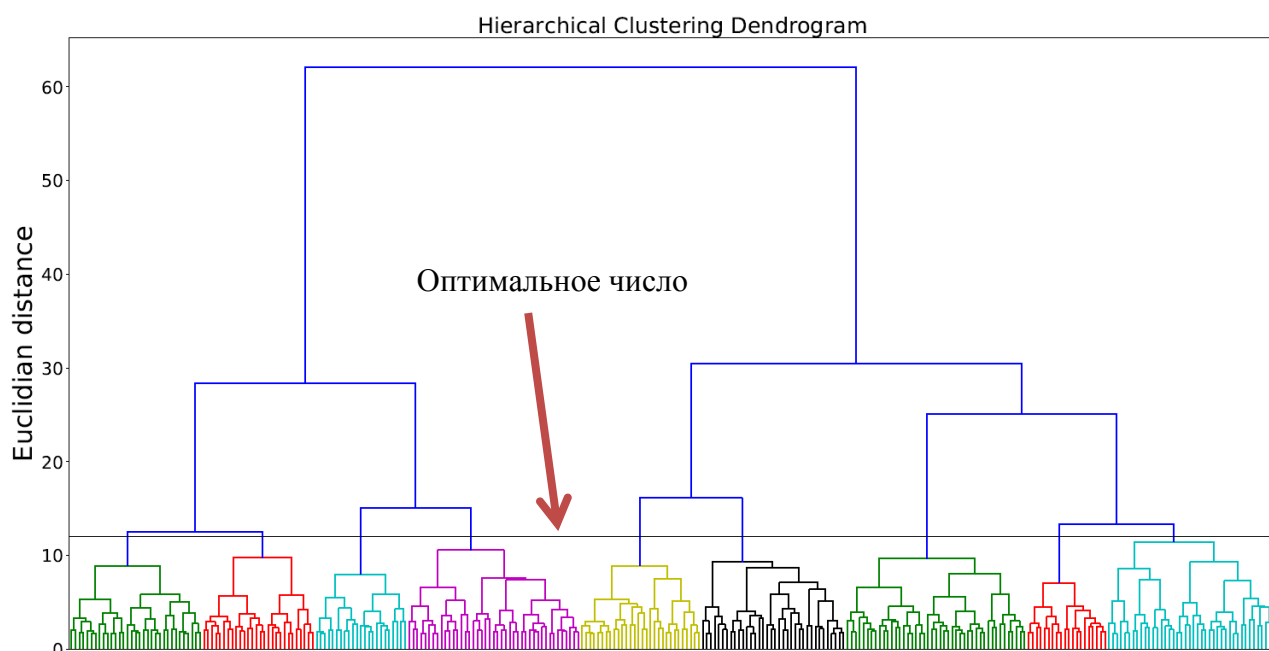


Рисунок 22 – Дендрограмма агломеративной кластеризации

Следует отметить, что только последние 300 слияний представлены на дендрограмме для того, чтобы сделать ее более презентабельной (изначальное количество кластеров – 3149 для 20 отобранных скважин). Оптимальным количеством кластеров является число девять. Девять кластеров также отмечены на дендрограмме при евклидовом расстоянии, равном 12. В работе были проверены все четыре типа агломеративной кластеризации, наиболее показательные результаты позволил получить тип Ward при евклидовом расстоянии. Другие виды расстояний не позволили получить лучшие результаты [8,12].

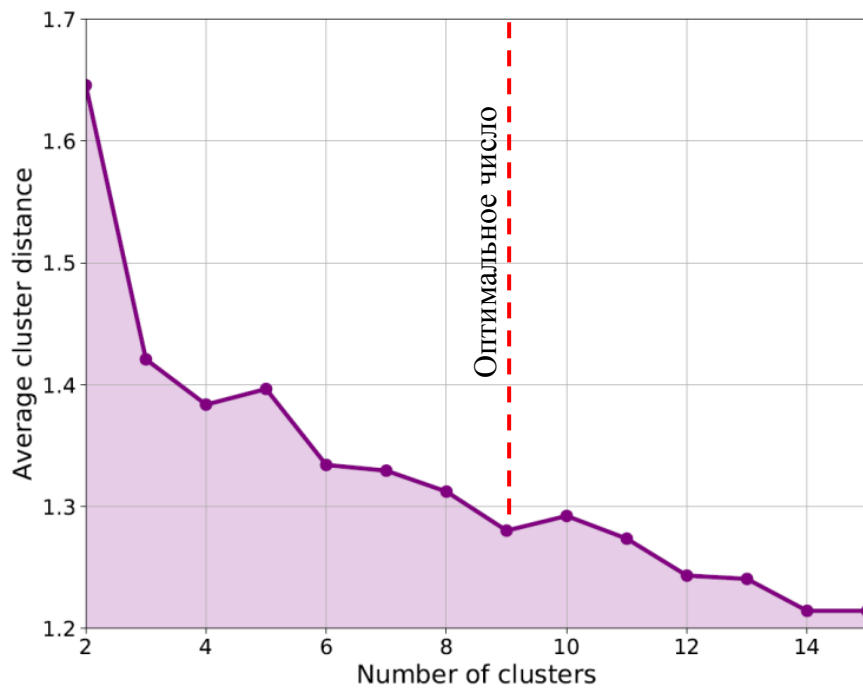


Рисунок 23 – Оценка агломеративной кластеризации

Более того, также была применена математическая оценка, которая аналогично показала девять кластеров как оптимальное значение.

После оптимизации параметров алгоритмов, они были применены на выборке из 20 скважин при выбранных параметрах. Результаты представлены на Рисунке 24 и Рисунке 25.

Очень важным преимуществом машинного обучения без учителя является выделение пропластков неколлектора там, где человек выделил коллектор. Например, в скважине D84 видно, что нижнее песчаное тело, выделенное человеком как массивный песчаник, машина поделила на части, выделив в нем пропластки неколлектора. На Рисунках 24, 25 левая колонка – результат интерпретации ГИС, правые четыре колонки – результаты кластеризации четырьмя алгоритмами. Кластеры серого цвета были отнесены к неколлекторским, тогда как цветные кластеры – коллекторские. В результате можно сделать вывод о том, что зачастую машина может получить результаты лучше, чем может получить человек ввиду его ограниченности способностей.

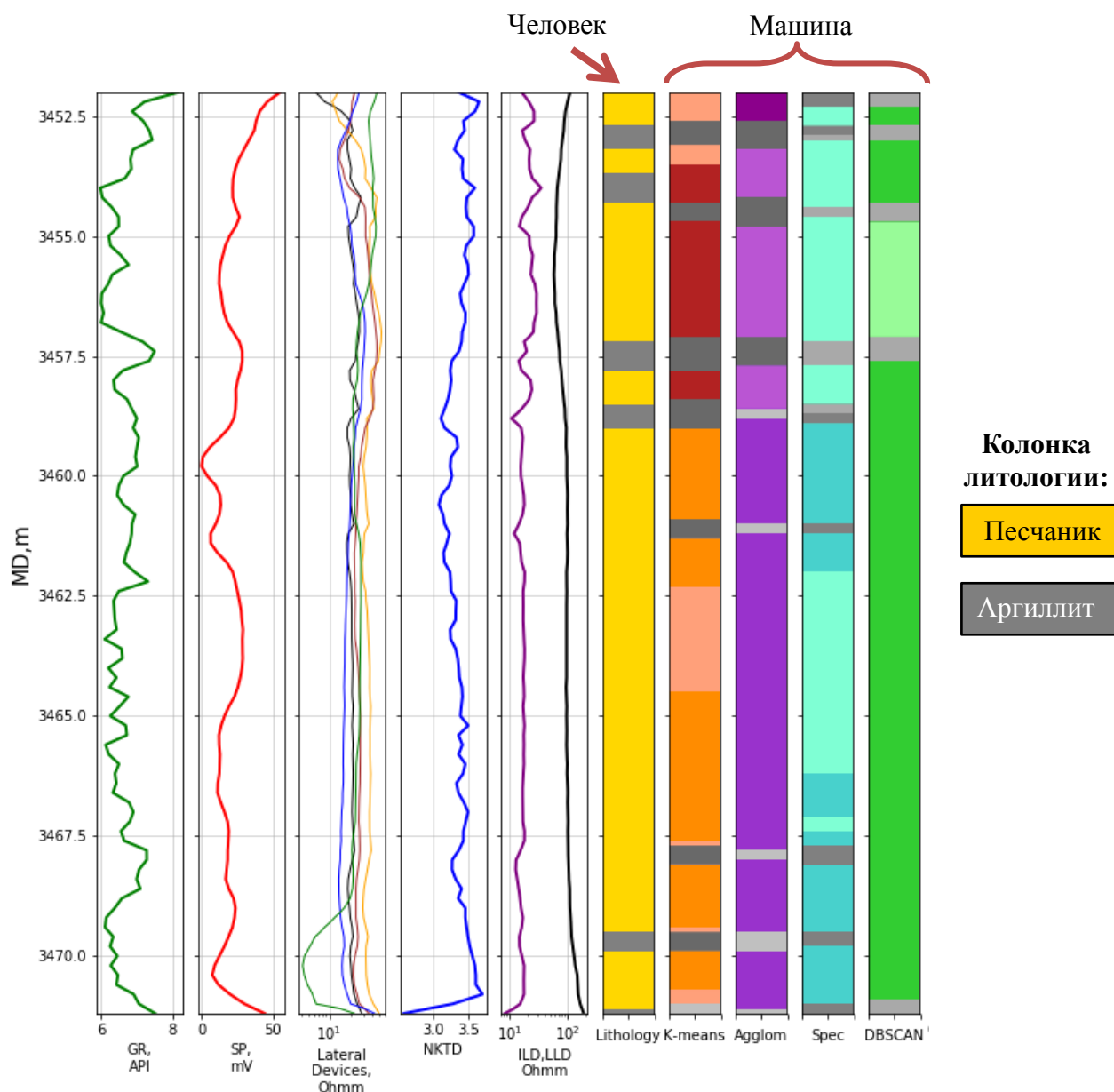


Рисунок 24 – Кластеризации на скважине D84

Более того, видно, что алгоритм K-means отражает разбиение на группы в каротажах сильнее, тогда как DBSCAN позволяет получить очень плохие результаты, так как, например, вообще не различает пропластки неколлектора. В то же время спектральная кластеризация не может различить породы также хорошо, как K-means и агломеративная кластеризация, так как один и то же кластер был выделен в средней и верхней частях скважины D84.

В результате мы приходим к выводу, что лучшим алгоритмом для конкретной геологической задачи для скважины D84 является K-means.

Для остальных 19 скважин он также хорошо работает, например, в скважине D169 он просто лучше остальных сходится с колонкой РИГИС.

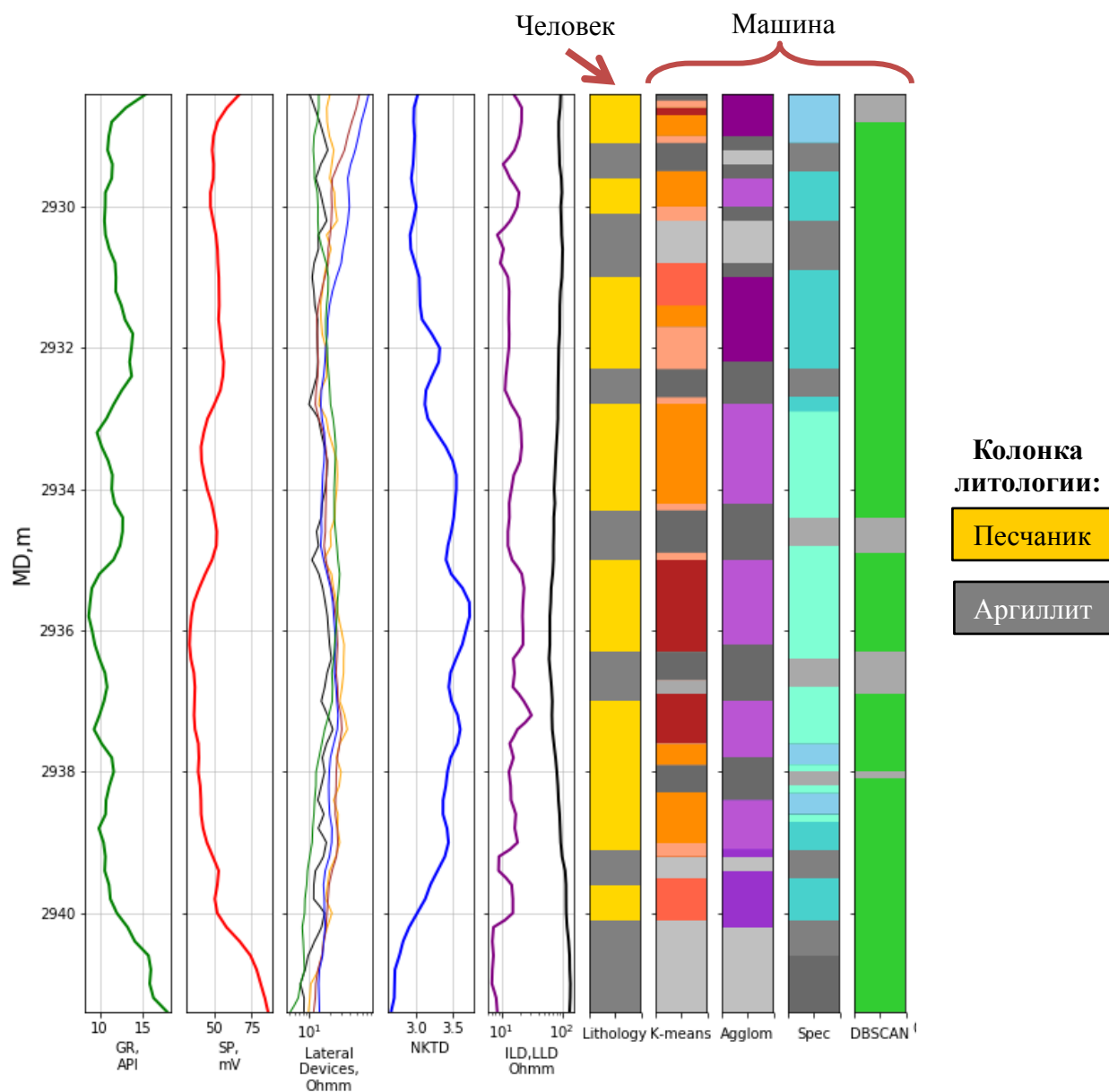
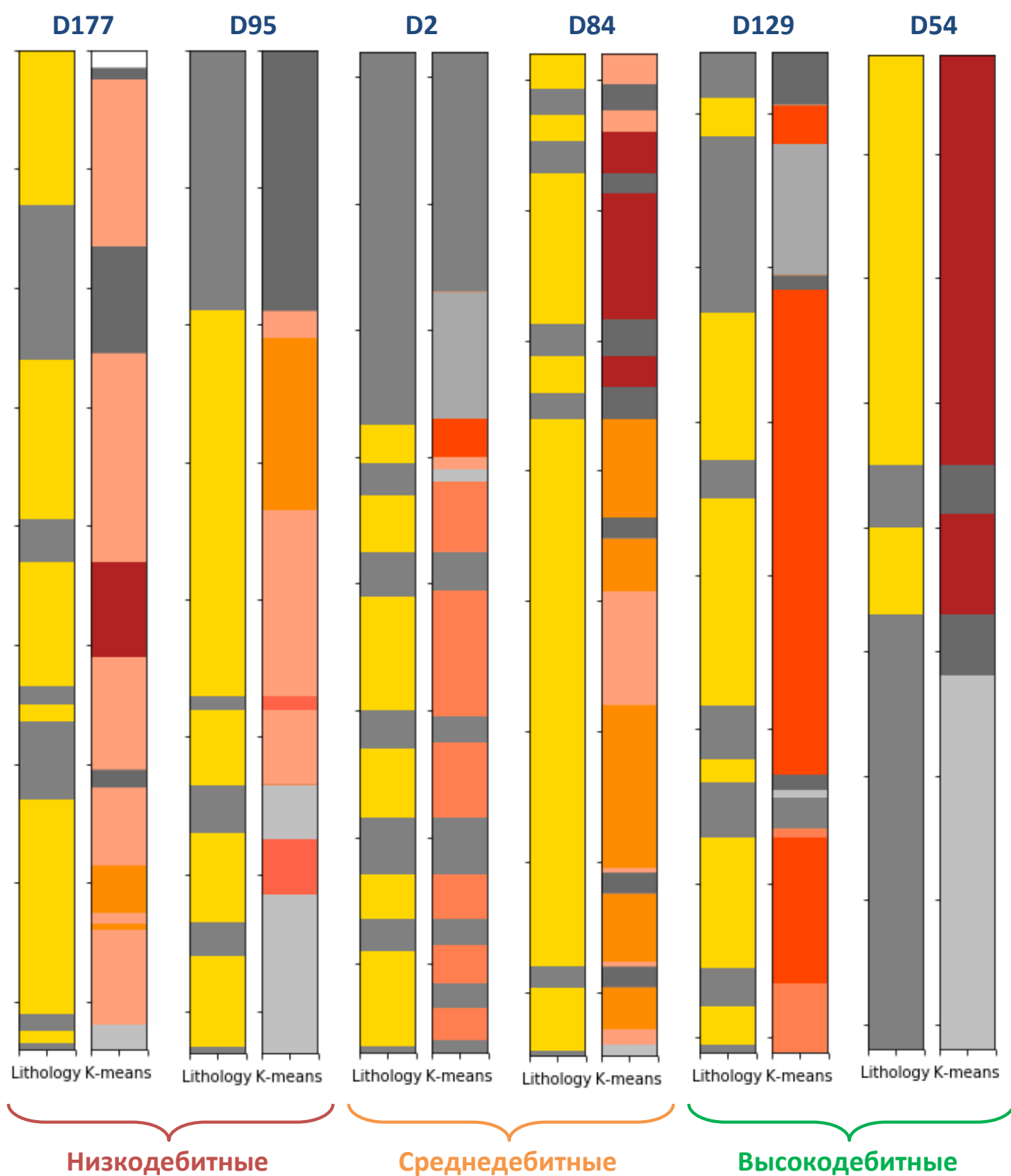


Рисунок 25 - Кластеризация на скважине D169



K-means кластеризация:



Где **NR** – неколлекторские кластеры, **R** – коллекторские кластеры

Рисунок 26 – Результаты кластеризации K-means для низкодебитных, среднедебитных и высокодебитных скважин

Таблица 2 – Промышленные скважинные характеристики

Скважина	Запускной дебит нефти, т/сут	Начальная обводненность	Тип скважины
D177	10	0.79	Низкодебитная/ Высокая обводненность
D95	12	0.68	
D2	33	0.35	Среднедебитная/ Средняя обводненность
D84	42	0.28	
D129	57	0.25	Высокодебитная/ Низкая обводненность
D54	94	0.18	

Необходима интерпретация результатов кластеризации для того, чтобы понять, насколько хорошо совокупность кластеров и их мощности характеризуют промышленные характеристики скважин, так как в задачу входит не только уверенное выделение продуктивных интервалов, но и понимание того, насколько сильно промышленные характеристики зависят от этих кластеров, например, запускные дебиты и обводненность. Из Рисунка 26 видно, что высокодебитные скважины вскрывают преимущественно кластеры R5 и R6. Анализируя эти кластеры по скважинам, было определено, что это песчаник с высокими значениями нефтенасыщенности (больше 0.6) с высокой средней проницаемостью. Конечно, проницаемость не может быть в полной мере отражена только в каротажах, так как не существует прямой зависимости проницаемости и любой каротажной кривой, поэтому проницаемость является лишь скрытой характеристикой, которая отражена в небольшой степени, но все-таки отражена. Что касается низкодебитных скважин с высокой начальной обводненностью, то они вскрывают в большей степени кластеры R1 и R2. Особенно кластер R1 показывает наиболее сильное влияние на низкие дебиты и высокую обводненность, это песчаник с низкой нефтенасыщенностью (0.35 – 0.45) с изменчивой средней проницаемостью по скважинам.

Таким образом, интерпретация кластеров показывает, что имеется связь между кластерами и промышленными характеристиками скважин.

Следующим шагом является применение алгоритма с выбранными оптимальными параметрами для скважин всего месторождения.

Некоторые скважины были исключены из анализа, например, горизонтальные скважины, боковые стволы и те, в которых отсутствуют важные каротажи. Оставшиеся 149 скважин использовались при общей кластеризации.

После того, как данные были подготовлены, скважины были обработаны алгоритмом кластеризации K-means, после чего мощности каждого кластера в каждой скважине были оценены. Соответствующие карты мощностей кластеров представлены на Рисунках 27-29. При построении карт был использован метод Кригинг для интерполяции мощностей в межскважинное пространство. Все скважины впоследствии также были визуально оценены на сходимость с РИГИС, из чего был сделан вывод о высокой точности выделения алгоритмом продуктивных интервалов.

Анализируя карты толщин кластеров можно сделать вывод о том, что «хорошие» кластеры, которые являются нефтенасыщенными, а скважины в этих кластерах дают высокие запускные дебиты, расположены в большей степени в ЧНЗ, в то время как «плохие» кластеры расположены по краям ЧНЗ, в водонефтяной зоне и за пределами ВНК. Кластеры неколлектора распространены повсеместно.

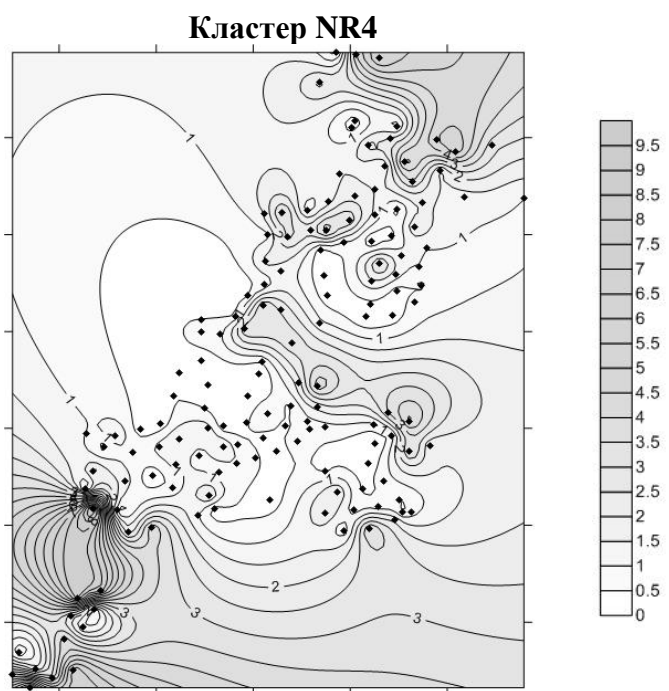
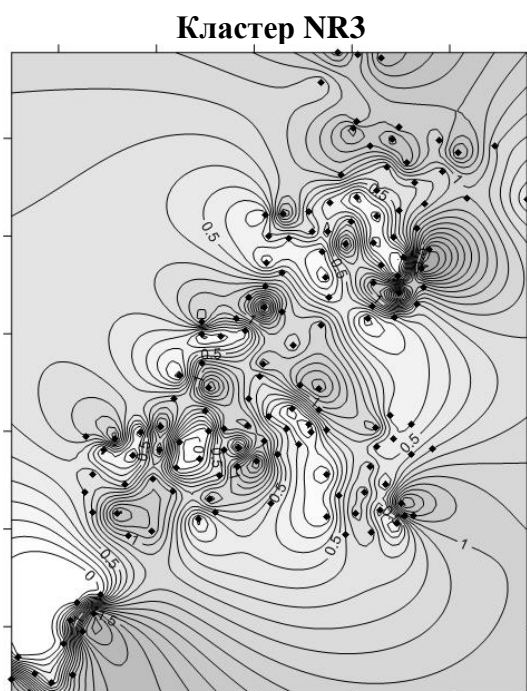
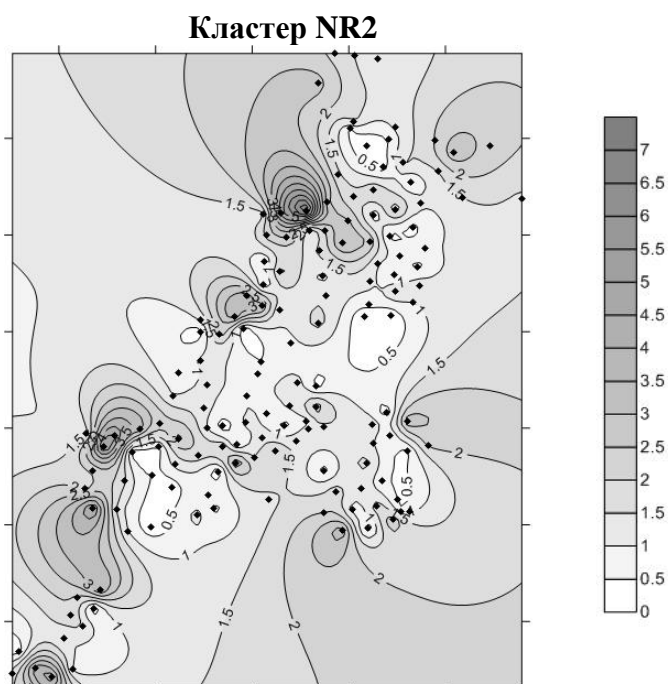
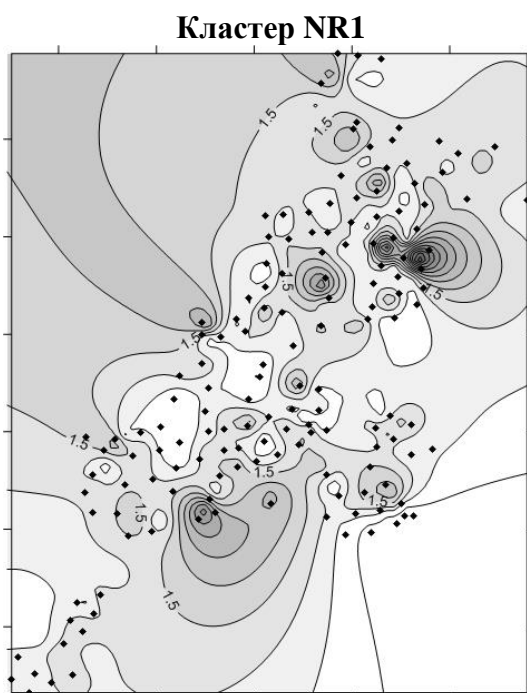


Рисунок 27 – Карты толщин неколлекторских кластеров

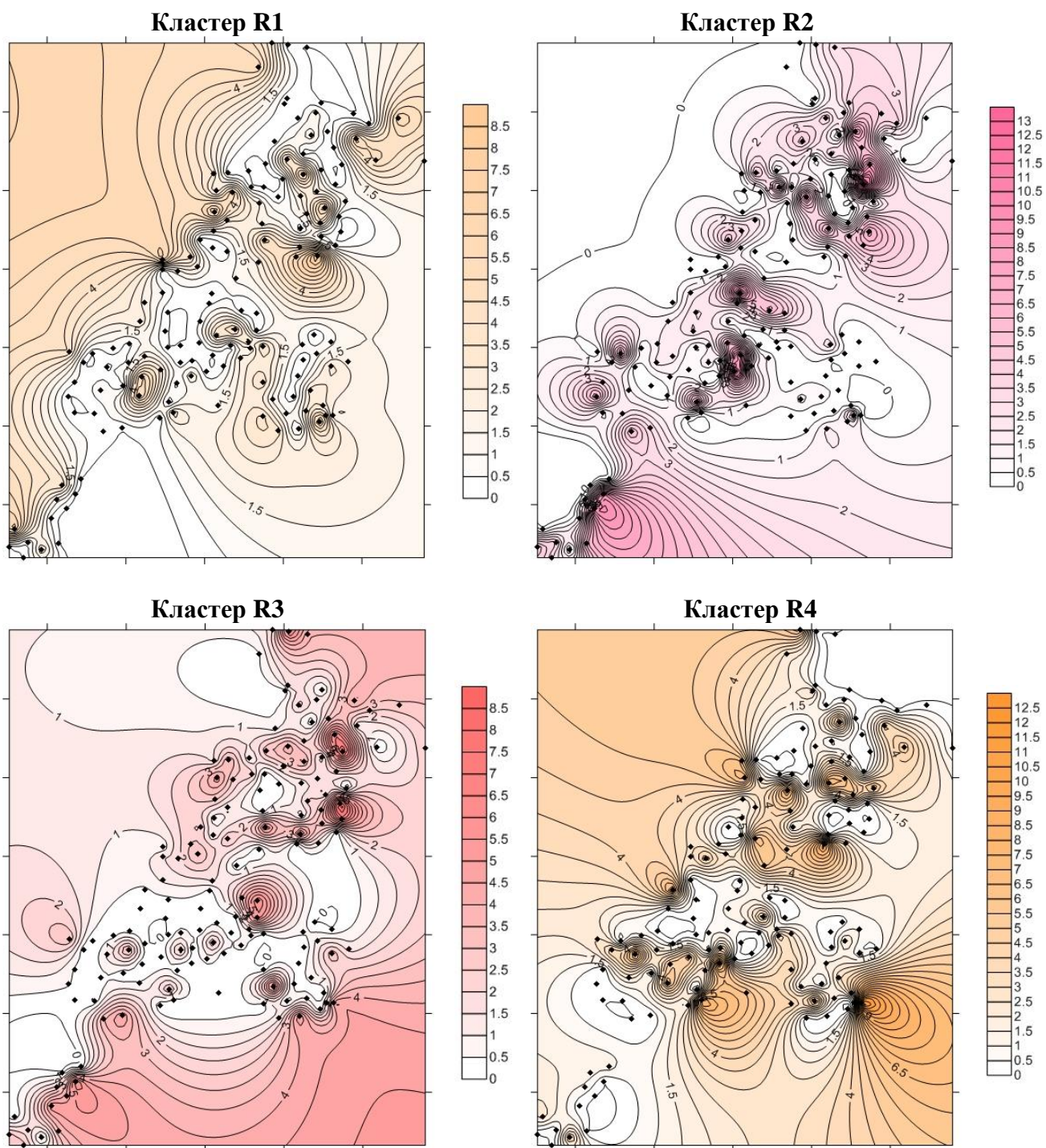


Рисунок 28 – Карты толщин коллекторских кластеров

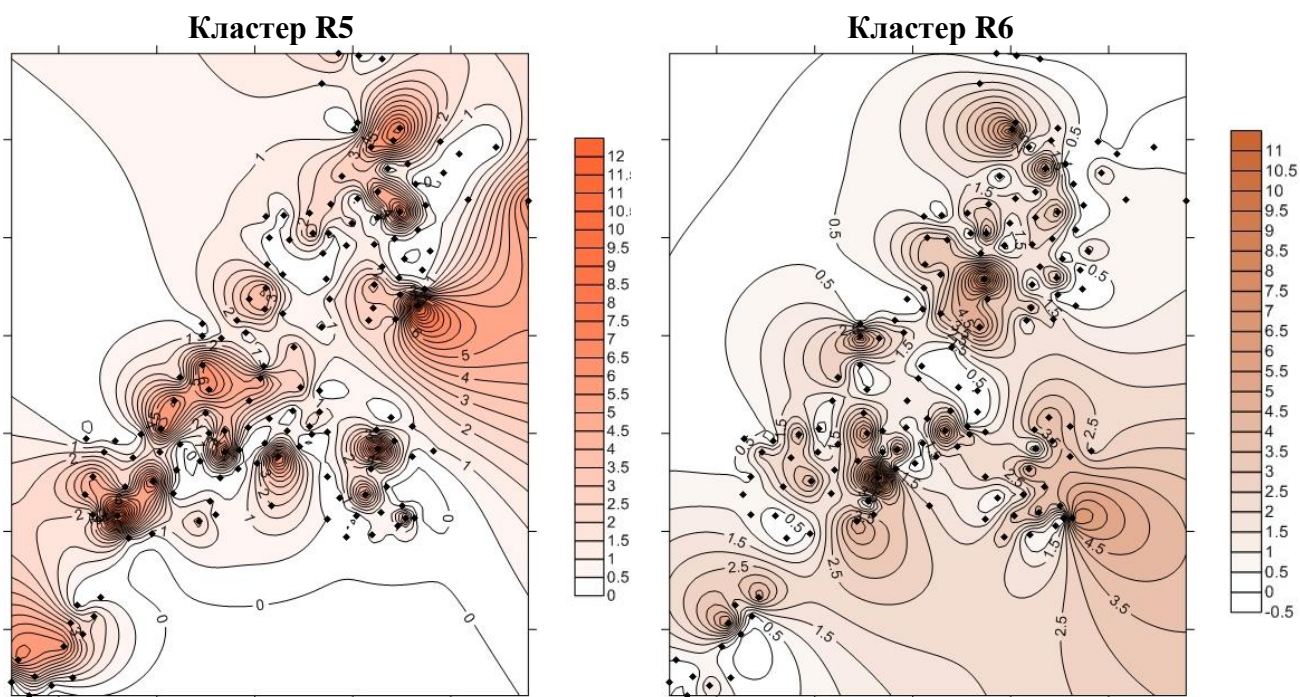


Рисунок 29 – Карты толщин коллекторских кластеров

4.5 Машинное обучение с учителем для прогноза запускных дебитов

Важной целью является не только использование кластеризации для геологических целей, но также результаты, полученные в процессе кластерного анализа, которые требуется увязать с промышленными скважинными характеристиками, такими как запускные дебиты. Раздел предсказаний запускных дебитов можно отнести к решению проблемы регрессионного анализа, которая является частью раздела машинного обучения с учителем и позволяет предсказывать непрерывные данные.

Ранее было описано, что на качественном уровне прослеживается связь типов кластеров и их мощностей в скважинах с промышленными характеристиками. Поэтому требуется дальнейшая проработка этой идеи с целью получения количественного результата по участкам месторождения, на которых планируется бурение новых скважин.

Прогноз добычи должен быть основан не только на геологических данных, таких как эффективная толщина и типы кластеров, но и на динамических параметрах, таких как, например, давление, скин-фактор.

Проблема регрессии является поиском линии аппроксимации, линии тренда, которая может быть описана математически и в дальнейшем

использована для получения совершенно новых независимых данных. То есть, после того, как модель определена и натренирована, что значит, что, получена необходимая линия тренда, описываемая переменной, которая предсказывается, может быть получена путем подстановки описывающих переменных в математическое уравнение линии тренда [21].

В конкретной задаче, которую требуется решить в работе, необходимо определиться с описывающими переменными, которые могут влиять на описываемый результат. То есть, описываемыми переменными являются запускной дебит по нефти и жидкости. Процедура получения модели регрессии представлена на Рисунке 30 [21].



Рисунок 30 – процесс использования машинное обучение с учителем для предсказания непрерывных данных

В первую очередь, для оценки описывающих переменных, требуется проверка их на мультиколлинеарность. Под мультиколлинеарностью понимается такой эффект, при котором описывающие переменные зависят друг от друга. В таком случае может возникнуть проблема, если эти переменные привносят большой вклад в описание описываемой переменной, то есть имеют большой вес. При наличии двух зависящих переменных с большим весом, занижается важность остальных переменных, а результат будет зависеть, в конечном счете, по сути, от одной переменной, что считается плохим результатом [21]. При нахождении взаимозависимых переменных, одну из них рекомендуется исключить из анализа [21].

Мультиколлинеарность проверяется нахождением коэффициентов корреляции попарно. В качестве коэффициента корреляции обычно используют коэффициент Пирсона. Он показывает, насколько хорошая прослеживается линейная связь между двумя переменными. Коэффициент может иметь отрицательные значения, что показывает обратную корреляцию. Идеальная линейная корреляция описывается значением коэффициента 1. Полное отсутствие корреляции – 0, Идеальная обратная корреляция описывается значением минус 1 [21].

$$r = \frac{\sum_{i=1}^n [(x^{(i)} - \mu_x)(y^{(i)} - \mu_y)]}{\sqrt{\sum_{i=1}^n (x^{(i)} - \mu_x)^2} \sqrt{\sum_{i=1}^n (y^{(i)} - \mu_y)^2}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (3)$$

Где μ – среднее значения переменной, σ_{xy} - ковариация между двумя переменными, σ_x и σ_y – стандартные отклонения переменных.

Таким образом, анализируя коэффициенты корреляции, основной интерес представляют значения, меньше 0.8 или больше минус 0.8. Иные значения указывают на очень высокую корреляцию переменных, и одну из них следует исключить. Также, пристальное внимание следует уделять переменным в диапазоне 0.5 и 0.8, так как это тоже указывает на высокую корреляцию, и при возможности, следует избавляться от таких переменных, однако при необходимости, такие данные все же могут быть использованы [8,12,21].

Хорошим тоном принято считать построение матрицы корреляции между всеми параметрами, которая показывает полную картину взаимной корреляции. Выбранные для анализа описывающие переменные: мощности всех десяти кластеров в нормированном виде (нормированы на общую мощность), депрессия (dP) как динамическая характеристика месторождения, и стратиграфическая мощность пласта. Описываемыми переменными являются запускные дебиты нефти (O_rate) и жидкости (L_rate).

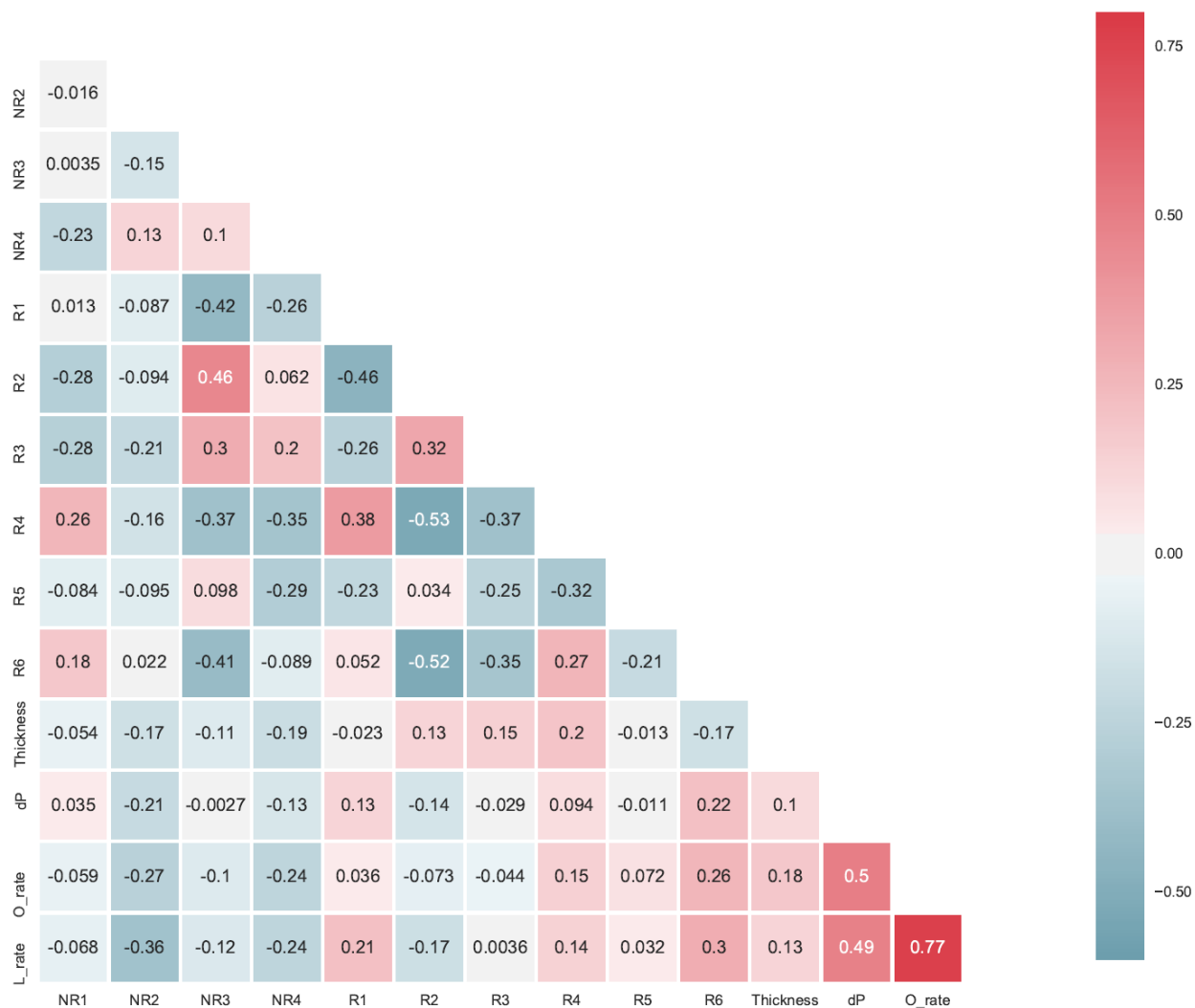


Рисунок 31 – Матрица корреляции

Анализируя матрицу корреляции, можно сделать вывод о том, что дебиты нефти и жидкости в большей степени зависят от депрессии и мощности кластера R6, что является логичным, и было выявлено на качественном уровне ранее. Помимо них, также сильное влияние оказывают мощности кластеров R5 и R4 [8,12,21].

Несмотря на то, что корреляция между параметрами довольно низкая, регрессия позволяет получить довольно высокий коэффициент корреляции, используя все параметры вместе. В нашем примере все коэффициенты корреляции ниже 0.5, что говорит о том, что исключение параметров не требуется, а все они могут быть использованы для получения уравнения аппроксимации [8,12,21].

Более того, набор попарных кросс плотов также может быть построен, где каждый кросс плот – график, показывающий зависимость между двумя параметрами. Набор таких кросс плотов представлен в Приложении А [8,12,21].

Регрессионный анализ, также как и кластерный, должен быть адекватно оценен разными регрессионными моделями, а затем должна быть выбрана оптимальная модель. Существуют основные виды регрессионных моделей, такие как линейная регрессия, лассо, градиентный бустинг, и ридж регрессия. Одна из них должна быть выбрана для предсказания дебитов.

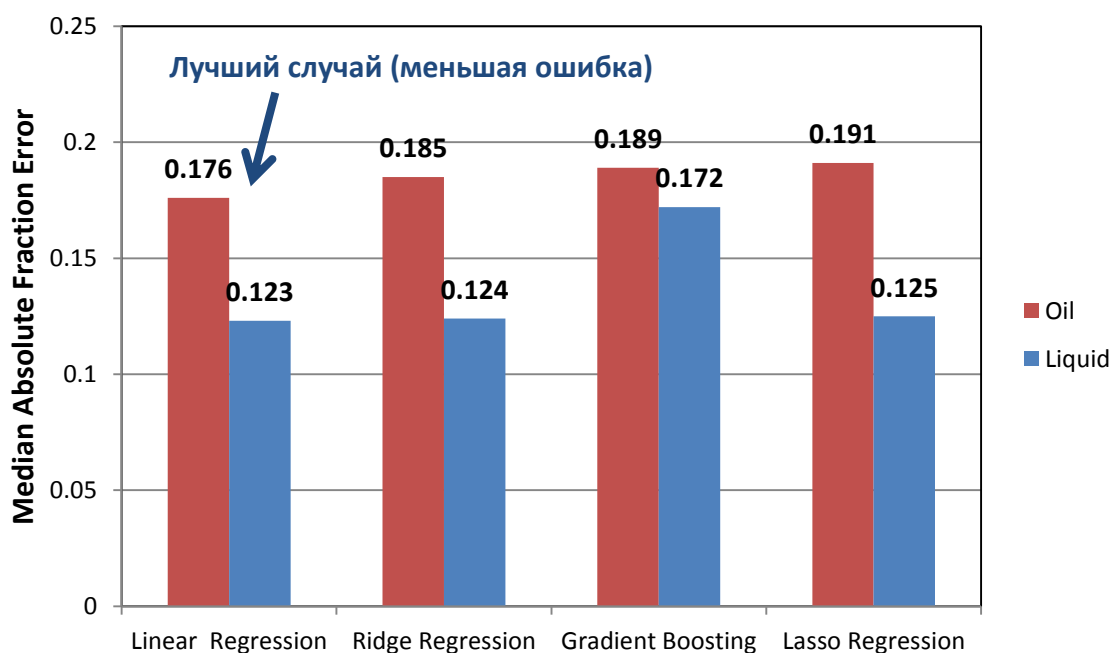


Рисунок 32 – Выбор модели регрессии

Лучшей моделью является модель линейной регрессии, так как она имеет наименьшую медианную абсолютную ошибку для прогноза запускного дебита нефти и жидкости. Более того, существенным плюсом линейной регрессии является то, что ее можно использовать где угодно, получив коэффициенты, так как это просто уравнение прямой линии в многомерном пространстве, имеющее пересечение линии и углы наклона в каждой плоскости проекции [8,12,21].

Качество прогноза оценивается с помощью кросс-валидации, например, K-fold кросс-валидации. Каждый раз, когда модель проходит процедуру тренировки, необходимо разделить весь набор данных на тренировочную и

тестовую выборки. Тренировочная выборка используется для настройки модели, а тестовая – для ее оценки на новых данных, которые не участвовали в тренировке. Так как очень важно, чтобы значения тренировочной и тестовой выборок не сильно различались, необходимо итеративно делить все данные на тестовую и тренировочную выборки, чтобы каждое наблюдение было и в виде тренировочной выборки, и в виде тестовой. Этот процесс и называется кросс-валидацией, он позволяет избежать возможных проблем с выбросами в данных [8,12,21].

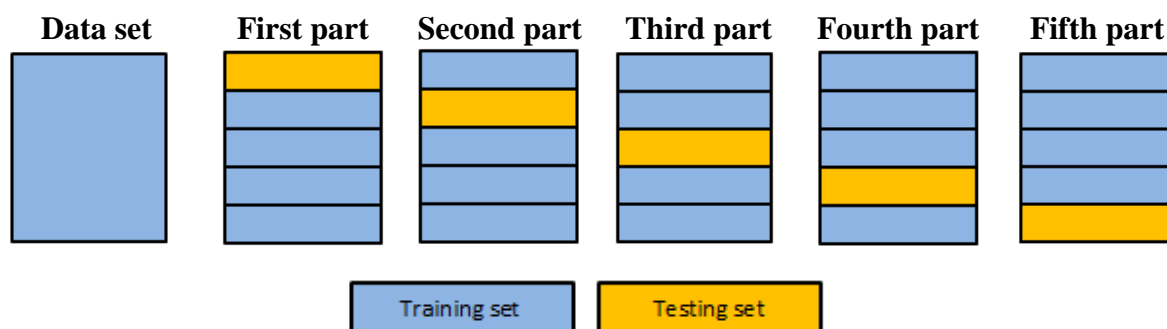


Рисунок 33 – K-fold кросс-валидация (K=5)

K-fold кросс-валидация были применена для всех типов моделей регрессии, количество частей, на которые был поделен набор данных – 20 частей, так как данных немного, и необходимо использовать в каждой итерации максимально возможное количество наблюдений в виде тренировочной выборки. То есть, в каждой итерации 141 наблюдение являлось тренировочной выборкой, а 8 – тестовой. Результат построение кросс - плота для кросс-валидации представлен на Рисунках 34 и 35. Кросс плоты для остальных моделей представлены в Приложении Б [8,12,21].

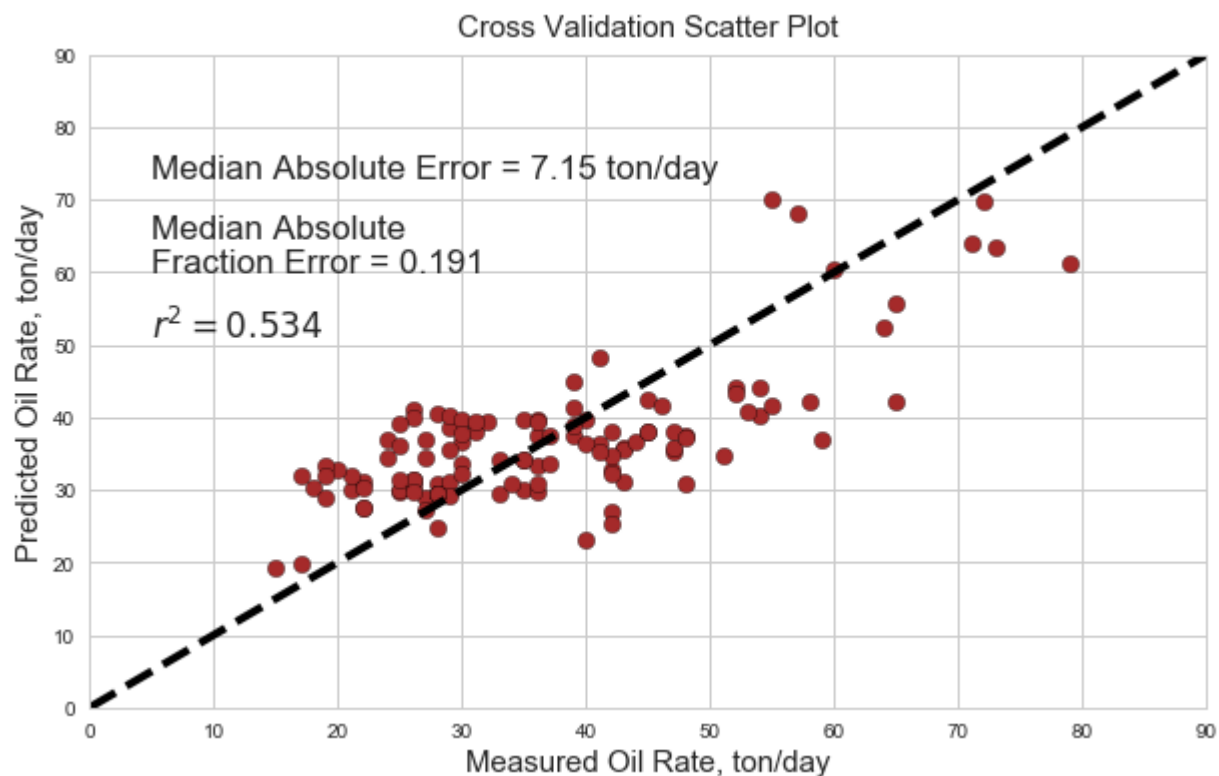


Рисунок 34 – Оценка модели прогноза запускного дебита нефти с помощью кросс-валидации

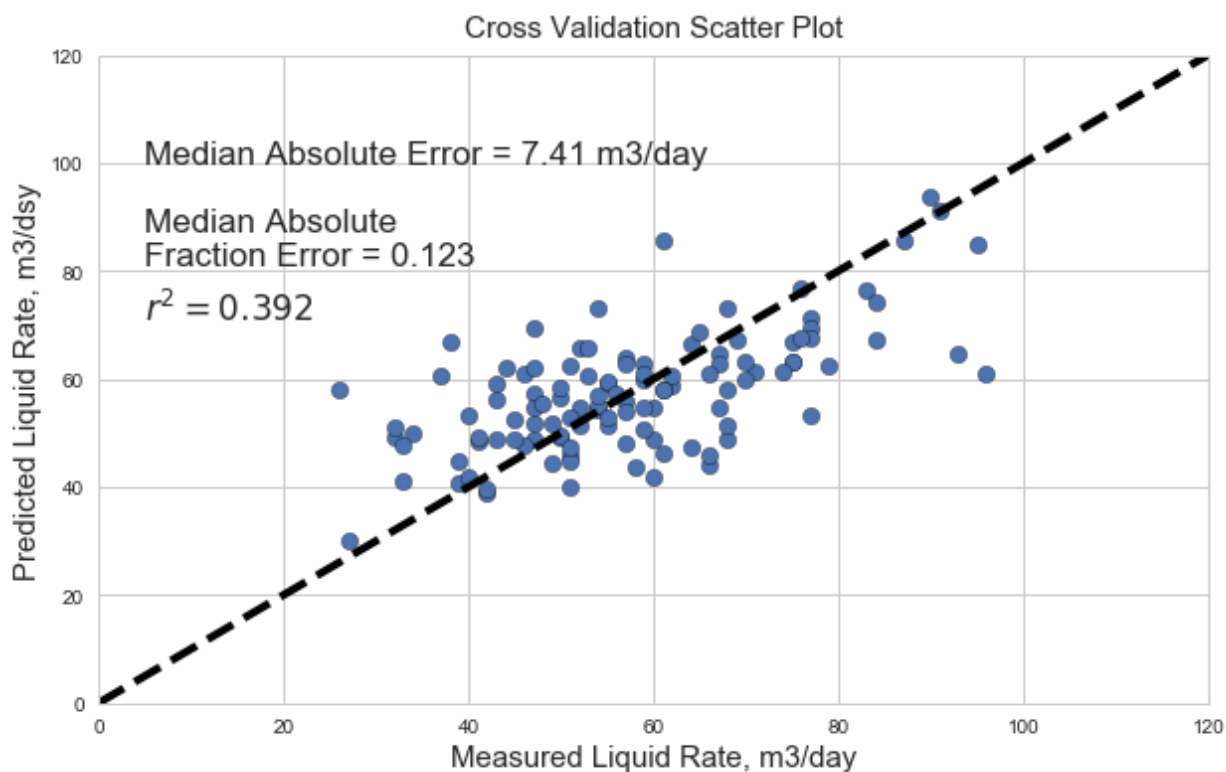


Рисунок 35 – Оценка модели прогноза запускного дебита жидкости с помощью кросс-валидации

При оценке моделей использовался параметр медианной абсолютной ошибки вместо r^2 , так как r^2 может давать высокие значения, если большая часть точек лежит в районе линии единичного наклона, притом что часть значений в интервале высокодебитных и низкодебитных скважин совершенно не предсказывается. Задачей является получение модели с низкой ошибкой во всем интервале запускных дебитов. В конечном счете, можно сказать, что модели не являются максимально хорошими ввиду того, что не был использован скин-фактор, так как данные о нем не были предоставлены, а расчет был невозможен. Для улучшения модели в дальнейшем рекомендуется использовать также значения скин - фактора [8,12,21].

Модель линейной регрессии – линейное уравнение, в котором предсказываемый параметр описывается описывающими переменными с определенными весами (коэффициентами):

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 \dots + a_nx_n, \quad (4)$$

где a_0 – проекция пересечения линии на ось описываемой переменной, $a_1 - a_n$ – углы наклона проекции линии на соответствующие плоскости.

В нашей конкретной задаче уравнение приобретает вид:

$$Q_o \left(\frac{\text{ТОН}}{\text{СУТ}} \right) = a_0 + a_1 \cdot \text{NR1} + a_2 \cdot \text{NR2} + a_3 \cdot \text{NR3} + a_4 \cdot \text{NR4} + a_5 \cdot \text{R1} + a_6 \cdot \text{R2} + a_7 \cdot \text{R3} + a_8 \cdot \text{R4} + a_9 \cdot \text{R5} + a_{10} \cdot \text{R6} + a_{11} \cdot \text{Мощность(м)} + a_{12} \cdot \text{dP(атм)}, \quad (5)$$

$$Q_l \left(\frac{\text{м}^3}{\text{СУТ}} \right) = b_0 + b_1 \cdot \text{NR1} + b_2 \cdot \text{NR2} + b_3 \cdot \text{NR3} + b_4 \cdot \text{NR4} + b_5 \cdot \text{R1} + b_6 \cdot \text{R2} + b_7 \cdot \text{R3} + b_8 \cdot \text{R4} + b_9 \cdot \text{R5} + b_{10} \cdot \text{R6} + b_{11} \cdot \text{Мощность(м)} + b_{12} \cdot \text{dP(атм)}, \quad (6)$$

Где NR1-NR4, R1-R6 – нормализованные мощности кластеров, мощность в метрах, депрессия в атмосферах.

Таблица 3 – Коэффициенты модели линейной регрессии

Коэффици- ент	Значе- ние	Коэффици- ент	Значе- ние	Коэффици- ент	Значе- ние	Коэффици- ент	Значе- ние
a₀	14.913	a₇	14.100	b₀	38.136	b₇	20.082
a₁	-22.416	a₈	22.500	b₁	-31.087	b₈	6.990
a₂	-18.786	a₉	15.032	b₂	-43.564	b₉	11.510
a₃	-40.095	a₁₀	24.466	b₃	-8.185	b₁₀	34.551
a₄	-8.292	a₁₁	-0.604	b₄	-14.249	b₁₁	-0.508
a₅	-3.658	a₁₂	0.177	b₅	19.379	b₁₂	0.162
a₆	17.148	-	-	b₆	4.574	-	

5. РЕЗУЛЬТАТЫ РАЗРАБОТКИ

В процессе работы была разработана методика, которая может быть применена на разных стадиях разработки месторождения. В первую очередь, был обоснован алгоритм процесса выделения продуктивных интервалов в скважинах с помощью кластеризации, который может быть применен как для построения геологической модели, так и для проверки интерпретации ГИС. Основными решениями в этой части работы явилась разработка универсальной метрики оценки качества кластеризации, которую рекомендуется применять в схожих задачах, связанных с кластерным анализом. Также детально описан процесс подготовки данных для кластеризации, обоснование количества кластеров, обоснование и выбор лучшего алгоритма, и дальнейший процесс кластеризации всего фонда скважин.

После, была найдена связь между результатом кластеризации и промышленными характеристиками скважин, что в дальнейшем сподвигло на доработку методики для конкретной задачи составления рейтинга бурения, используя результаты кластерного анализа. Для решения этой задачи было применено машинное обучение с учителем для получения регрессионной модели прогноза запускных дебитов скважин, как по жидкости, так и по нефти. Оценка экономических показателей проводилась по полученным профилям добычи проектных скважин, которые были построены с помощью метода кривых темпов падения добычи.

В конечном счете, в работе был получен результат, при котором выделение продуктивных интервалов в скважинах с помощью кластерного анализа происходит более детально и качественно, чем это делает квалифицированный интерпретатор.

Что касается составления рейтинга бурения, то при сравнении разработанной методики со стандартным подходом, который используют в нефтегазовой отрасли, результат получился более предсказуемым и разнородным, а дальнейшая экономическая оценка показала огромный эффект при использовании такого подхода в сравнении со стандартным.

6. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ

6.1 Перспективность методики

Разработанная методика, позволяющая выделять продуктивные интервалы достаточно быстро и уверенно с помощью алгоритмов машинного обучения без учителя, является перспективной с точки зрения уменьшения времени затрачиваемого на интерпретацию геофизических исследований. Таким образом, при необходимости быстрой и качественной оценки эффективных толщин в скважинах, на которых интерпретация ГИС проведена не была по разным причинам (новая скважина или недостоверные результаты интерпретации) применение методики рекомендуется. Более того, кластерный анализ позволяет не только определять эффективные толщины в скважинах, но и находить скрытые зависимости в каротажных кривых и разделять коллектор на подтипы в соответствии с геологической обстановкой, фильтрационно-емкостными свойствами и насыщением. Такая проблема особенно актуальна для карбонатных коллекторов, где без поиска скрытых зависимостей, зачастую, интерпретация ГИС затруднительна. В результате метод позволяет экономить время работников нефтегазовых компаний в случае с необходимостью быстрой оценки пробуренных новых скважин, а также может являться дополнительным инструментом для подтверждения интерпретации ГИС или нахождения ошибок при работе со «сложными коллекторами» [14].

Более того, при необходимости, метод может быть расширен до задачи составления рейтинга бурения при условии использования машинного обучения с учителем и дальнейшего анализа темпов падения скважин на месторождении. Основной проблемой на этом шаге является решение задачи регрессии. На этом круг применения методики не заканчивается, так как геологическая основа кластерного анализа позволяет проводить фациальное зонирование месторождения, что неизбежно приведет к более детальным и точным концептуальным и геологическим моделям [14].

6.2 Реализация технической задачи

Использование методики на данном этапе предполагает знание пользователем языка программирования Python, а также достаточно углубленных знаний машинного обучения. В перспективе возможна оптимизация процесса кластерного анализа путем разработки комплексного программного обеспечения в виде плагина для Petrel, Schlumberger [14]. Это позволит миновать длительный процесс подготовки данных для анализа, так как все необходимые данные уже загружены в программу при построении геологической модели. Более того, так как карты толщин кластеров достаточно легко могут быть построены в Petrel, проблема с выгрузкой и загрузкой данных в разных пакетах программного обеспечения будет минимизирована, а процесс построения карт может быть автоматизирован. Таким образом, это избавит пользователя от необходимости владения языком программирования, что позволит использовать методику специалистам любых специальностей [14].

Также задача составления рейтинга бурения может быть упрощена внедрением основных шагов в программное обеспечение «NGT Smart» и «РН КИН».

7. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ

Так как методика, разработанная и представленная в работе, несет сугубо теоретический характер, ее следует применять для достижения максимально эффективного процесса разработки месторождения и экономии времени и средств, то основными нормами и обязанностями в связи с ней являются меры, направленные на снижение проблем и экологических загрязнений. Источники загрязнения и возможных проблем, аварийных ситуаций связаны с бурением скважин и их последующей эксплуатацией.

В процессе бурения скважин следует использовать оборудование и инструмент, прошедший сертификацию, оборудование должно иметь соответствующие сертификаты и паспорта [3].

Оборудование не может использоваться по истечению полезного срока действия, так как это может повлечь за собой возможные аварии ввиду поломки оборудования [3].

При использовании оборудования, которое может травмировать персонал, следует устанавливать специальные отражатели и ограждения с целью минимизации доступа персонала к работающему оборудованию. Ограждающее оборудование должно быть окрашено в специальные предупреждающие цвета и нести предупреждающую информацию на себе [3].

Что касается самого процесса установки оборудования и бурения, то оборудование должно быть установлено на прочных фундаментных основаниях и опробовано в соответствии с инструкцией изготовителя [3].

По завершении монтирования буровой вышки и сопутствующего оборудования, вся конструкция должна быть проверена в динамике без нагрузки на крюк для выявления возможных допущенных ошибок при монтировании. При выявлении нарушений работы оборудования и поломок требуется немедленно устранить их до начала работ, повторный запуск оборудования запрещается [3].

Только после завершения монтажа оборудования и подписания акта о вводе в эксплуатацию происходит подача электричества на оборудование. Подача должна осуществляться в светлое время суток. Также на вышке должны быть установлены источники аварийного освещения с автономным питанием. Обязательным требованием является установка источников освещения во взрывозащитном исполнении [3].

Установка шланга для нагнетания промывочной жидкости требует нанесения специальной обмотки металлическим канатом диаметром 12 мм. Это требуется для избегания ущерба здоровью и жизни персонала при нарушении герметичности шланга [3].

Запуск ротора может быть осуществлен только после снятия колонны с клиновых захватов и при отсутствии другого оборудования на нем.

Обязательное условие осуществления бурения – установка противовыбросового оборудования в соответствии с допустимым максимальным давлением в процессе бурения на устье скважины. Оборудование должно иметь необходимый запас по давлению, чтобы в том числе была возможность его использования по окончании бурения при освоении скважин. Оборудование должно периодически проверяться на закрытие и открытие. При недолжном срабатывании оборудования необходима обязательная замена оборудования. Продолжение бурения с неисправной системой противовыбросового оборудования запрещено [3].

При остановке оборудования на длительное время требуется проверка системы на наличие взрывоопасных веществ и удаление последних до начала работ.

Во время бурения обязательным условием является контроль следующих параметров бурения:

1. вес колонны на крюке;
2. реологические свойства бурового раствора, его плотность;
3. расход раствора на входе и выходе и их сравнение;
4. контроль давления на манифольде.

При расхождении расходов на входе и выходе, увеличении уровня раствора в емкостях требуется немедленная остановка процесса бурения и замер давления в колонне и манифольде. При подтверждении притока пластовой жидкости в скважины необходимо провести расчет количества поступившего флюида, его тип, пластовое давление и необходимую плотность бурового раствора для глушения скважины, после чего производится процесс глушения замещением бурового раствора на новый [3].

Все обязательства по обеспечению безопасного бурения возлагаются на бурового мастера и буровую бригаду на срок проведения буровых работ [3].

При эксплуатации скважин необходима установка лубрикаторов на фонтанной арматуре. Все операции на скважине необходимо проводить только через лубрикатор с закрытой центральной задвижкой. Проведение операций с открытой скважиной запрещены во всех случаях, когда иной возможности проведения операций нет. Если по каким-то причинам приходится снимать фонтанную арматуру (например, замена НТК), то обязательным условием является установка противовыбросового оборудования на скважине. Во время остановки работ на скважине требуется закрывать противовыбросовое оборудование [4].

На скважинах, оборудованных насосными установками, должна быть предусмотрено дистанционное отключение. Установка электроцентробежных насосов требует особых компетенций при работе со скважиной, так как к скважине подается кабель высокого напряжения [4].

Запрещается запуск насоса при отсутствии жидкости в нем, так как это приведет к его выходу из строя. При работе с ШГНУ необходим контроль динамограммы с целью своевременного выявления проблем работы насоса [4].

Все электродвигатели, установленные в пределах кустовой площадки, должны быть заземлены.

7.1 Аварийные ситуации в процессе бурения и добычи

Основными аварийными ситуациями являются:

1. допущение открытого фонтанирования, как в процессе бурения, так и в процессе капитального ремонта скважин;
2. взрывы горючих веществ;
3. утечки химически-опасных веществ;
4. обрыв электрического кабеля, электрическое замыкание.

При допущении открытого фонтанирования при бурении требуется немедленное закрытие противовыбросового оборудования. Это же касается и проведения капитального ремонта скважин. При отсутствии возможности остановки фонтанирования средствами установленного оборудования, по возможности требуется произвести ликвидацию оборудования из зоны аварии для возможности беспрепятственного доступа бригады ликвидации ГНВП. Всех сотрудников требуется эвакуировать из зоны аварии. При обрыве кабеля требуется незамедлительно уведомить начальство о проблеме, как и в случае с любыми другими авариями. Трогать и приближаться к кабелю, тем более пытаться самостоятельно устранить проблему категорически запрещается. При работе с источниками электричества обязательным условием является ношение специальной защитной одежды, выполненной из диэлектрических материалов [3,4].

Для того чтобы минимизировать риск аварийных ситуаций требуется:

1. необходимый мониторинг и контроль всех систем в соответствии с установленными требованиями безопасности;
2. необходимый контроль всех источников возможного возникновения неисправностей и возгораний;
3. настройка звуковых индикаторов неисправностей и поломок всех систем;

Основной рекомендацией для ведения производства является исполнение всеми без исключения работниками правил безаварийной работы и охраны окружающей среды [3,4].

7.2 Охрана окружающей среды

Производственный процесс несет за собой риски загрязнения окружающей среды. Вред природе наносится с первого дня ведения работ на территории месторождения. В первую очередь, когда происходит отсыпка кустов и строительство дорог, в окружающую среду выбрасываются выхлопные газы от тяжелой техники. Также происходит вырубка лесов для отсыпки кустов. Происходит загрязнение почв, воды. Таким образом, основными экосистемами, которые страдают от загрязнения, являются атмосфера, биосфера, гидросфера и литосфера [1].

Что касается атмосферы, то основные источники загрязнения могут быть поделены на две группы: органические и неорганические источники загрязнения. К органическим источникам загрязнения относятся такие, как испарение воды в процессе работы, оставшиеся после работ металлические отходы (трубы и прочее), системы аэрации. Из неорганических источников могут быть выделены лакокрасочные работы, двигатели внутреннего сгорания и выделяющиеся в результате их работы в атмосферу выхлопные газы. Для снижения загрязнений на производстве планируется проводить сбор всего мусора и вывоз его на полигон, что позволит избежать органических источников загрязнения. Также большой вред атмосфере наносят пары и углекислый газ, выделяющиеся в процессе сжигания попутного газа на газопоршневых установках [1].

В результате работ также страдает гидросфера, как поверхностные, так и подземные воды. В процессе производства планируется отбор подземной воды для питьевых нужд персонала, а также отбор воды из сеноманского горизонта для закачки воды в систему ППД. Также в результате аварий могут страдать наземные воды при разливе нефти. Для сохранения гидросферы планируется сбор использованных вод в специальные емкости и вывоз их на полигон. Сброс вод в наземные водоемы, а также закачка их в пласты строго запрещена. Для борьбы с возможными разливами нефти на территории кустовых площадок

планируется установка обваловок, которые помогут избежать загрязнения водоемов.

Более того, в процессе ведения работ страдает и литосфера. Наиболее частым источником загрязнения при разработке месторождений является загрязнение почв как в процессе строительства оборудования и отсыпки кустовых площадок, так и в результате разливов нефти, бурового раствора. Для сохранения почв обязательным условием является установка обваловок, а также применение специальных амбаров для сброса бурового шлама и раствора. Впоследствии шлам и использованный буровой раствор вывозится на полигон.

Основным источником вреда биосфере является необходимость вырубки лесов для обустройства месторождения. Также в результате возможных разливов нефти может страдать и животный мир. Для ограничения представителей фауны от процесса ведения деятельности требуется установка отгораживающих заборов на кустовых площадках, а также в местах нахождения людей. Своевременное удаление источников загрязнений позволит обезопасить животных от контакта с загрязняющими веществами. Также планируется высадка лесов для восстановления понесенного ущерба флоре [1].

7.3 Ликвидация скважин

Большой ущерб может нанести некачественная ликвидация месторождения. Ликвидация необходима по окончании разработки месторождения, когда ликвидируется и вывозится все оборудование с территории. Для осуществления этого процесса в экономическом расчете учтены затраты на ликвидацию скважин в размере 10% от затрат на строительство.

В процессе ликвидации обязательным условием является снятие всего оборудования как с установок подготовки нефти и газа, так и с кустовых площадок, а именно, фонтанных арматур, шлейфов, трубопроводов.

Для качественной и безопасной ликвидации скважин требуется глушение скважин, их цементирование до устья и установка бетонного блока, изолирующего устье скважины и не дающее доступа к нему.

7.4 Безопасность и здоровье персонала

От безопасного производства зависит уровень работников, так как квалифицированный персонал не имеет желания работать на производстве, где велик риск получения травм. Более того, компания считает своим долгом обустройства на территории месторождения безопасного производства. В рамках безопасного производства планируется:

1. проведение ежемесячных обязательных медицинских осмотров, без которого запрещается приступать к работе;

2. установка ограждающих приспособлений для ограничения доступа персонала к установкам, могущим нанести вред здоровью человека;

3. рабочие условия персонала должны быть спланированы в соответствии с требованиями законодательства по необходимому достаточному освещению, вентилированию помещений, особенно в местах возможных утечек нефти и газа. Обязательные сбор и вывоз мусора с территории месторождения. Более того, в связи с тем, что территория месторождения расположена в достаточно жестких погодных условиях, особенно в зимнее время, требуется обязательное поддержание достаточной для комфортной работы температуры в помещениях, а при работе на улице – выдача персоналу качественной теплой одежды на разное время года [5].

4. каждый сотрудник обязан сдавать ежегодные экзамены на наличие знаний об охране окружающей среды, медицинские тесты и оказание первой медицинской помощи;

5. на территории месторождения должны находиться медицинские работники разных специальностей, а также в имуществе компании обязательно должен быть вертолет для быстрой транспортировки пострадавших до ближайших населенных пунктов для оказания медицинской помощи.

ЗАКЛЮЧЕНИЕ

В данной работе была разработана методика, позволяющая выделять продуктивные интервалы с помощью алгоритмов машинного обучения без учителя, а затем применять полученные при кластерном анализе результаты для прогноза промышленных характеристик скважин с помощью регрессионного анализа.

В работе были решены следующие задачи:

1. Проведен анализ геологического строения выбранного месторождения, исходных данных в виде ГИС и РИГИС, а также данных о добычи скважин;
2. Был разработан подход для обработки исходных необходимых для работы данных, а также проведена ее автоматизация с помощью языка программирования Python, составлена программа, автоматизирующая процесс;
3. Был разработан алгоритм для работы с машинным обучением без учителя в области нефтяной индустрии, была разработана универсальная метрика, позволяющая оценить качество кластеризации любого алгоритма, особенно хорошо показавшая себя в геологической задаче, проведен процесс кластеризации и обоснован выбор оптимального алгоритма для кластерного анализа с помощью разработанной метрики;
4. Был составлен план обработки результатов кластеризации, в результате которой формируются данные, используемые в дальнейшем регрессионном анализе;
5. Описан процесс выбора регрессионной модели для конкретной геологической задачи для прогноза запускных дебитов нефти и жидкости на скважинах;
6. Обоснована возможность улучшения системы разработки путем бурения уплотняющих скважин, спроектированы уплотняющие скважины на месторождении, получены типы и мощности кластеров в проектных скважинах с применением методов интерполяции;

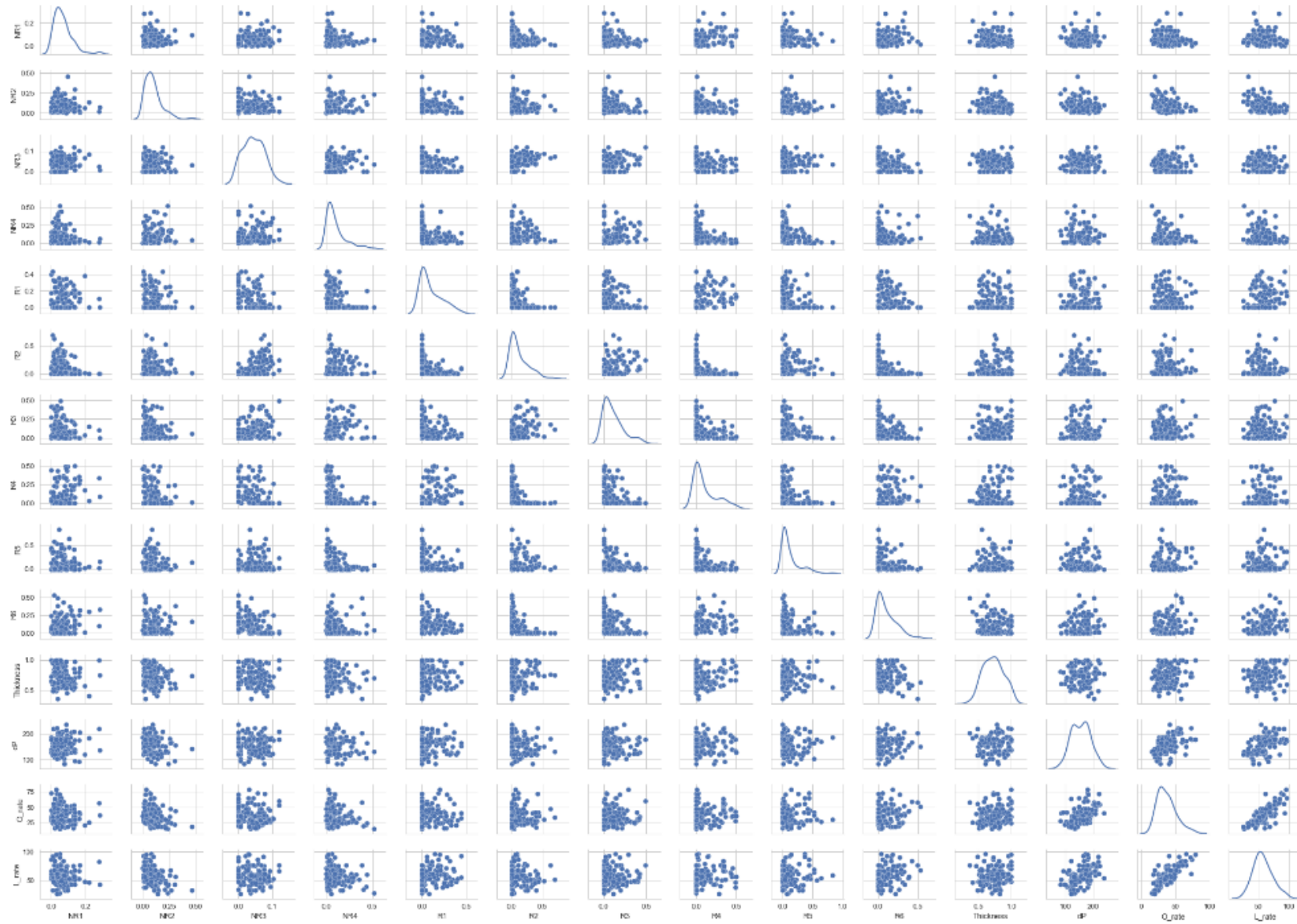
7. Получены запускные дебиты на проектных скважинах с помощью регрессионной модели;
8. Проведен анализ кривых темпов падения нефти и жидкости, получены усредненные кривые падения добычи нефти и жидкости для гиперболической модели Арпса;
9. Проведена оценка методики при составлении рейтинга бурения проектных скважин путем сравнения ее со стандартным подходом;
10. Получена экономическая оценка методики;
11. Проведен анализ возможных негативных последствий на окружающую среду, которые могут проявиться в процессе бурения уплотняющих скважин, и описаны способы их минимизации, также были спланированы мероприятия по обеспечению безопасности рабочих на территории месторождения.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Федеральный закон от 10 января 2002 г. №7-ФЗ «Об охране окружающей среды» // Собрание законодательства Российской Федерации - 14 января 2002 г. - №2 - Ст. 133.
2. Налоговый кодекс Российской Федерации. [Электронный ресурс]: от 13.07.2015 № 214-ФЗ, №232-ФЗ (ред. от 10.10.2016) // Консультант Плюс. - Режим доступа: <http://www.consultant.ru/>
3. Приказ от 12 марта 2013 года N101 «Об утверждении Федеральных норм и правил в области промышленной безопасности "Правила безопасности в нефтяной и газовой промышленности" (с изменениями на 12 января 2015 года) (редакция, действующая с 1 января 2017 года)»// Федеральная служба по экологическому, технологическому и атомному надзору
4. Приказ Министерства труда и социальной защиты РФ от 10 марта 2015 г. N 153н "Об утверждении профессионального стандарта "Бурильщик капитального ремонта скважин" // Собрание законодательства Российской Федерации - 2013, N 4, ст. 293; 2014, N 39, ст. 5266
5. Приказ от 9 декабря 2014 г. N 997н «Об утверждении типовых норм бесплатной выдачи специальной одежды, специальной обуви и других средств индивидуальной защиты работникам сквозных профессий и должностей всех видов экономической деятельности, занятым на работах с вредными и (или) опасными условиями труда, а также на работах, выполняемых в особых температурных условиях или связанных с загрязнением» // Министерство труда и социальной защиты российской федерации
6. Доусон, М. Прографируем на python – Санкт-Петербург: Питер, 2018. – 416 с.
7. Кныш, С.К. Основы структурной, исторической и региональной геологии / С.К. Кныш, Н.В. Гумеров, А.К. Полиенко. – Томск: Изд-во ТПУ, 2009. – 375 с.
8. Рашка, С. Python и машинное обучение / перевод с английского А.В. Логунова. – Москва: ДМК Пресс, 2017. – 418 с.

9. Чернова, О.С. Нефтегазоносные провинции России и сопредельных территорий / О.С. Чернова. – Томск: Изд-во ТПУ, 2008. – 256 с.
10. Hedong Sun Advanced Production Decline Analysis and Application / Hedong Sun. - Houston: Gulf Professional Publishing, 2015. – 352 p.
11. Patrick Corbett MSc Petroleum Geoscience / Patrick Corbett, Gary Couples, Andy Gardiner, Helen Lever. – Edinburgh: Heriot Watt University UK, 2017. – 640 p.
12. Sebastian Rashka Python machine learning / Sebastian Rashka – Birmingham: Packt Publishing Ltd., 2015. – 454 p.
13. Zed A. Shaw Learn Python the hard way: a very simple introduction to the terrifying beautiful world of computers and code – Third edition / Zed A. Shaw – Crawfordsville: RR Donnelley, 2013. – 306 p.
14. Y. Zee Ma Unconventional Oil and Gas Resources Handbook / Y. Zee Ma, Stephen A. Holditch. - Houston: Gulf Professional Publishing, 2015. – 550 p.
15. MSc Petroleum Geology – Aberdeen: University of Aberdeen UK, Department of Geology and Petroleum Geology, 2000. – 376 p.
16. MSc Petroleum Economics– Edinburgh: Heriot Watt University UK, Institute of Petroleum Engineering, 2016. – 556 p.
17. Joel Gevirtz Lithofacies Classification for Earth Modeling: Letting the Data Speak For Itself / Joel Gevirtz, Adriana Ovalle. SPE-185049-MS, 2017
18. Свободная энциклопедия Petrowiki, «Production forecasting decline curve analysis»
19. Электронный ресурс Habr, «Первичный анализ данных с Pandas»
20. Электронный ресурс Habr, «Обучение без учителя: PCA, кластеризация»
21. Электронный ресурс Habr, «Линейные модели классификации и регрессии»

ПРИЛОЖЕНИЕ А. МАТРИЦА КРОСС - ПЛОТОВ



ПРИЛОЖЕНИЕ Б. РЕЗУЛЬТАТ ОЦЕНКИ КРОСС-ВАЛИДАЦИЕЙ

